SEQUENCE CHARACTERISTICS OF BLADDER CANCER

CROSS-REFERENCE TO RELATED APPLICATIONS

The present application claims the benefit of priority to PCT application PCT/US00/41005, filed September 27, 2000, which claims the benefit of priority under 35 U.S.C. §119(e) of U.S. Provisional Patent Application Number 60/156,153, filed September 27, 1999, both of which are incorporated herein by reference.

BACKGROUND OF THE INVENTION

FIELD OF THE INVENTION

The present invention relates to the identification of polynucleotide sequences that are differentially expressed in bladder cancer. More specifically, the present invention relates to the use of the sequences and gene products for diagnosis and as probes.

DESCRIPTION OF RELATED ART

Bladder cancer is the second most-common genitourinary cancer in the United States, with only prostate cancer being more frequently diagnosed. Bladder cancer accounts for approximately two percent of all malignant tumors and approximately seven percent of all urinary tract malignancies in U.S. men. Over 54,000 new cases were estimated to be diagnosed in the United States in 1998, with approximately 12,500 deaths predicted [American Cancer Society, 1998]. The prevalence of bladder cancer is higher in industrialized nations, perhaps reflecting increased exposure to environmental carcinogens. Men are three times more frequently affected than women. The disease usually occurs between 60-70 years of age and the age-adjusted bladder cancer rate in white men is almost twice that of black men. Most bladder cancers (over 90%) are carcinomas of the transitional epithelium of the bladder's mucosal lining (transitional cell carcinoma (TCC)). Although 90 percent of the cases are localized at diagnosis, up to 80 percent recur.

A number of etiological factors are associated with the development of bladder cancer, but in industrialized countries, cigarette smoking is the most significant. Specific chemicals have also been identified as causing bladder cancer, as have a number of occupational exposures to less well-defined specific agents. Treatment with cytostatic drugs, especially cyclophosphamide, is associated with increased risk of bladder cancer, as is treatment with radiotherapy for uterine cancer.

Bladder cancer is a potentially preventable disease, with a significant morbidity and mortality in many parts of the world.

Tumors are graded according to the degree of cellular abnormality, with the most atypical cells being designated as high-grade (i.e., G3 grade) tumors. The major prognostic factors in carcinoma of the bladder are the depth of invasion into the bladder wall and the degree of differentiation of the tumor. The higher the grade of the tumor at the diagnosis, the higher the incidence of death from the disease within two years.

The stage of development of the tumor is significant in estimating disease prognosis. Most superficial, non-invasive tumors are papillary tumors which do not invade the *lamina propria*, and are classified as non-invasive TCC, i.e., "Ta" tumors, and they can recur, but nearly 70% will not progress further. A tumor which does not invade the muscle but does enter the *lamina propria* presents in many cases a worse prognosis. Such tumors are also classified as non-invasive TCC but are termed T1 tumors. Most superficial tumors are well differentiated and classified as G1 grade tumors. Patients in whom superficial tumors are less differentiated, large, multiple, or associated with carcinoma *in situ* in other areas of the bladder mucosa, (classified as G2-G3 tumors) are at greatest risk for recurrence and the development of invasive cancer. Invasive bladder tumors tend to spread rapidly to the regional lymph nodes and then into adjacent structures. Overall, the five-year survival rate of TCC is 76 percent for whites and 55 percent for blacks.

One of the management problems is the fact that carcinoma of the bladder is frequently multifocal. The entire bladder epithelium and the lining of the entire urothelial cell tract can undergo malignant change. After apparently successful treatment of a bladder lesion, new tumors can occur at the same site (recurrence) or in other urothelial cells in the bladder. Approximately 30 percent of bladder carcinomas present as multiple lesions at the time of initial diagnosis. The early diagnosis of bladder cancer is central to the effective treatment of TCC. Presently, the detection of bladder tumors relies on intravenous pyelogram or other contrast studies to rule out urothelial involvement in the kidneys or ureters, and invariably cystoscopy which remains the accepted standard for diagnosis of mucosal abnormalities. There are no presently reliable methods available to easily and specifically identify the presence of bladder cancer cells. A variety of new technologies and potential tumor markers are being studied in bladder cancer and some are being translated into clinical use.

It is important to realize that all available results of the diagnostic value of tumor markers do not allow firm clinical recommendations, but tests based on biomarkers undoubtedly influence the management of bladder cancer in the near future. Several new markers have been already identified and even approved for use (e.g. bladder tumor antigen (BTA) markers, NMP22, FDP). However, their clinical use is limited [Grossman, 1998], due to sensitivity and specificity problems in conjunction with cystoscopic examination.

Furthermore, due to the high rate of disease recurrence, follow-up of TCC patients is obligatory. There is a need to eliminate the invasive cytoscopy method of diagnosis and of follow-up and replace it with a reliable and non-invasive method of diagnosis.

Approximately 70-80 percent of patients with newly diagnosed bladder cancer present with superficial, non-invasive bladder tumors. Those who do are often curable. Tumor patients with deeply invasive disease can sometimes be cured by complete surgical removal of the bladder, irradiation, or a combination of modalities that include chemotherapy, however the five-year survival rate is less likely for such tumors. It is therefore of major importance to detect new tools that aid in both the initial early diagnosis and in follow-up of non-invasive TCC tumors.

Adverse prognostic features associated with a greater risk of disease progression include the presence of multiple aneuploid cell lines, nuclear p53 overexpression, and expression of the Lewis-x blood group antigen [Hudson and Herr, 1995; Lacombe et al., 1996]. It has been postulated that p53 can be useful for predicting the level of aggression of the tumor and to identify patients who will not benefit from chemotherapy. However, only a very small, select group of patients with invasive disease can benefit from this approach [Ozen, 1998]. Several treatment methods (i.e., transurethral surgery, intravesical medications, and cystectomy) have been used in the management of patients with superficial

5

tumors, and each method can be associated with five-year survival in 55-80 percent of patients treated [Hudson and Herr, 1995; Torti and Lum, 1984]. Invasive tumors that are confined to the bladder muscle on pathologic staging after radical cystectomy are associated with an approximately 75 percent, five-year progression-free rate of survival. Patients with more deeply invasive tumors (which are also usually less well differentiated) experience five-year survival rates of 20-40 percent following radical cystectomy. When the patient presents with a locally extensive tumor that invades pelvic viscera or with metastases to lymph nodes or distant sites, a five-year survival rate is uncommon, but considerable symptomatic palliation can still be achieved.

Surgery is the main treatment method. The extent of surgery is dependent on the pathological stage of the disease. Early disease is generally treated by intravesical chemotherapy and transurethral resection. Locally invasive disease can usually be managed only by radical cystectomy and urinary diversion. Definitive (curative) radiotherapy is generally reserved for bladder cancer patients who are not candidates for surgery. For superficial, low-grade disease, chemotherapy is applied intravesically (directly into the bladder) to concentrate the drug at the tumor site and eliminate any residual tumor mass after resection. Systemic chemotherapy can also be used to manage advanced bladder cancer; compete response rates of 30-50 percent have been reported. Single agent chemotherapy has demonstrated limited success.

However, even following surgery and resection of non-invasive TCC tumors, frequent follow-up is required (every 3 months) in both non-invasive and invasive cases.

It would therefore be useful to be able to identify early stage TCC in bladder cancer which has a significantly higher cure rate and generally does not require surgery. In addition, it would be useful to identify markers that can be employed for early diagnosis and follow-up of both non-invasive and invasive TCC, as an efficient and non-invasive alternative to cytoscopy.

SUMMARY OF THE INVENTION

According to the present invention, there is provided a method of diagnosing the presence of bladder cancer in a patient by analyzing a patient-derived sample for the presence of a least one expressed gene wherein the high level of expression of the gene is indicative of bladder cancer. Also provided by the present invention is a polynucleotide sequence whose expression is indicative of bladder cancer. A marker for bladder cancer is also provided. There are also provided methods of diagnosing bladder cancer by screening for the presence of at least one expressed gene wherein the presence of the expressed gene is indicative of bladder cancer. Methods of treating and regulating bladder cancer-associated pathology by administering to a patient a therapeutically effective amount of a chemical compound which inhibits a gene

7

comprising the nucleic acids sequences of the present invention are also provided.

DESCRIPTION OF THE INVENTION

According to the present invention, purified, isolated and cloned nucleic acid sequences associated with bladder cancer are provided. More specifically, the polynucleotides of the present invention are described in Tables 1, 2, and 5 and the corresponding sequences are set forth in Tables 3, 4 and 6 respectively.

When referring to bladder cancer, both invasive and noninvasive forms are included. Bladder cancers can also be referred to as transitional cell carcinomas or "TCC".

The present invention further provides a method of diagnosing the presence of bladder cancer in a patient, including the steps of analyzing a tissue sample from the patient for the presence of at least one expressed gene (up-regulated) wherein the mRNA from the expressed gene hybridizes to at least one of the sequences in Tables 1 or 2, with hybridization occurring under conditions sufficiently stringent to require at least 95% base pairing.

Further the present invention provides antibodies directed against the gene products of the sequences of the present invention. The antibodies can be either monoclonal, polyclonal or recombinant and be used in immunoassays as described in the Methods herein below.

8

By regulate or modulate or control is meant that the process is either induced or inhibited to the degree necessary to effect a change in the process and the associated disease state in the patient. Whether induction or inhibition is being contemplated is apparent from the process and disease being treated and is known to those skilled in the medical arts. The present invention identifies genes for gene therapy, diagnostics and therapeutics that have direct causal relationships between a disease and its related pathologies and up- or down-regulator (responder) genes. That is, the present invention is initiated by a physiological relationship between cause and effect.

The present invention identifies polynucleotides named in Tables 1 and 2, and set forth in Tables 3 and 4 respectively, that can be utilized diagnostically in bladder cancer. Polynucleotides named in Table 1 were found to match sequences in data banks and were newly found in the present application to be upregulated in TCC. The polynucleotides named in Table 2 are either genes with unknown protein product or of unknown genes. All the polynucleotides named in both Tables 1 and 2 were found to be associated with TCC relative to normal bladder samples. The polynucleotides named in Table 5 have their corresponding sequences set forth in Table 6, some of which are novel.

Where the sequences are partial sequences, they are markers or probes for genes that are regulated in bladder carcinoma. By "regulated" it is meant that the genes can be either upregulated or downregulated, depending upon the specific gene. In general these partial sequences are designated

"Expressed Sequence Tags" (ESTs) and are markers for the genes actually

expressed *in vivo* and are ascertained as described herein. Generally, ESTs

comprise DNA sequences corresponding to a portion of nuclear encoded mRNA.

The EST has a length that allows for PCR (polymerase chain reaction), use as a

hybridization probe and is a unique designation for the gene with which it

hybridizes (generally under conditions sufficiently stringent to require at least

95% base pairing).

For a detailed description and review of ESTs and their functional

utility see WO 93/00353 which is incorporated in its entirety by reference. WO

93/00353 further describes how the EST sequences can be used to identify the

transcribed genes. The Example herein also describes a method of identification.

The present invention also provides a method of diagnosing the presence of

bladder cancer in a patient, by the expression of at least one expressed gene

(up-regulated) identified by the polynucleotides of the present invention set forth

in Tables 1-6. Methods of identification of hybridization results can include, but

are not limited to, immunohistochemical staining of the tissue samples. Further

for identification of the gene, *in situ* hybridization, Southern blotting, single strand

conformational polymorphism (SSCP), restriction endonuclease fingerprinting

(REF), PCR amplification and DNA-chip analysis using nucleic acid sequence of

the present invention as probes/primers can be used.

The present invention further provides proteins encoded by the

identified genes. The present invention further provides antibodies directed

against these proteins. The present invention further provides transgenic animals

and cell lines carrying at least one expressible gene identified by the present invention. The present invention further provides knock-out eukaryotic organisms in which at least one nucleic acid sequences as identified by the probes of the present invention and prepared as described in the Methods.

The present invention provides a method of diagnosing bladder cancer, in particular TCC, in a subject which comprises determining in a sample from the subject the level of expression of at least one polypeptide –encoding polynucleotide, wherein a higher level of expression of the polynucleotide compared to the level of expression of the polynucleotide in a subject free of bladder cancer is indicative of bladder cancer, and wherein the polypeptide – encoding polynucleotide comprises a polynucleotide selected from the group consisting of

(a) the polynucleotides listed in Tables 3, 4 and 6;

(b) polynucleotides having sequences that differ from the polynucleotides in (a), without changing the polypeptide encoded thereby; and

(c) polynucleotides which are at least 70% homologous to the polynucleotides of (a).

In a preferred method of the invention, the analyzing step comprises using mRNA from the expressed gene to hybridize to at least one of the sequences in Tables 3, 4 and 6. In other preferred methods of the invention, the analyzing step comprises using RT-PCR technology or using a specific antibody to detect the presence of a polypeptide encoded by said polynucleotide.

11

The present invention also provides a method of diagnosing of diagnosing stage Ta or stage T1 in TCC, which comprises determining in a sample from the patient the level of expression of at least one polypeptide –encoding polynucleotide, wherein a higher level of expression of the polynucleotide compared to the level of expression of the polynucleotide in a patient free of bladder cancer is indicative of stage Ta or stage T1, and wherein the polypeptide –encoding polynucleotide comprises a polynucleotide selected from the group consisting of

(a) the polynucleotides listed in Tables 3, 4 and 6;

(b) polynucleotides having sequences that differ from the polynucleotides in (a), without changing the polypeptide encoded thereby; and

(c) polynucleotides which are at least 70% homologous to the polynucleotides of (a).

The present invention also provides isolated polynucleotides which comprise a polynucleotide selected from the group consisting of:

(a) the novel polynucleotides listed in Tables 4 and 6;

(b) polynucleotides having sequences that differ from the polynucleotides in (a), without changing the polypeptide encoded thereby; and

(c) polynucleotides which are at least 70% homologous to the polynucleotides of (a).

The present invention also provides such polynucleotides wherein the polynucleotide comprises a polynucleotide having at least 30, preferably at least 40, nucleotides from the polynucleotides described above.

The present invention also provides compositions comprising the isolated polynucleotides of the invention.

The present invention also provides an isolated polypeptide encoded by a polynucleotide, wherein the polynucleotide comprises a polynucleotide selected from the group consisting of:

(a) the polynucleotides listed in Tables 3, 4 and 6;

(b) polynucleotides having sequences that differ from the polynucleotides in (a), without changing the polypeptide encoded thereby; and

(c) polynucleotides which are at least 70% homologous to the polynucleotides of (a).

The present invention also provides such a polypeptide, wherein the polypeptide is a portion which retains the biological activity thereof or a polypeptide which is at least substantially homologous or identical thereto.

The present invention also provides a peptide, wherein the peptide is dominant negative peptide which competes with the biological activity of the polypeptide.

The present invention also provides an antibody which binds to a unique epitope of the polypeptide of the invention. The present invention also provides a method of diagnosing bladder cancer in a patient which comprises determining in a sample from the patient the level of expression of at least one polypeptide, wherein a higher level of polypeptide compared to the level of the polypeptide in a patient free of bladder cancer is indicative of bladder cancer.

13

The method includes using an antibody, preferably wherein the presence of more than one polypeptide is detected by using more than one such antibody. The present invention also provides a method of treating bladder cancer-associated pathology in a subject by administering to the subject a therapeutically effective amount of a chemical compound which inhibits a gene, or polypeptide encoded thereby, which comprises a polynucleotide selected from the group consisting of:

(a) the polynucleotides listed in Tables 3, 4 and 6;

(b) polynucleotides having sequences that differ from the polynucleotides in (a), without changing the polypeptide encoded thereby; and

(c) polynucleotides which are at least 70% homologous to the polynucleotides of (a).

The present invention also provides a gene therapy vehicle for delivering a polynucleotide of the invention to a subject, whereby the polynucleotide is expressed in the target cells of the subject. The present invention also provides isolated antisense oligonucleotides complementary to the polynucleotides of the invention.

The samples from the subjects which are used for diagnosis comprise samples of urine, blood, saliva, tissues and cells of all types; urine samples are preferred. A control sample includes a normal equivalent sample derived from a healthy subject.

The term "antibody" includes polyclonal antibody, single chain antibody, Fab fragment, monoclonal (MAB), polyclonal and recombinant

14

antibodies. A molecule which comprises the antigen-binding portion (CDR) of an antibody specific for a polypeptide, variant or fragment is also included in the term "antibody".

Negative dominant peptide refers to a partial cDNA sequence that encodes for a part of a protein, i.e. a peptide (see Herskowitz, 1987). This peptide can have a different function from the protein from which it was derived. It can interact with the full protein and inhibit its activity or it can interact with other proteins and inhibit their activity in response to the full protein. Negative dominant means that the peptide is able to overcome the natural proteins and fully inhibit their activity to give the cell a different characteristic, like resistance or sensitization to killing. For therapeutic intervention either the peptide itself is delivered as the active ingredient of a pharmaceutical composition or the cDNA can be delivered to the cell utilizing the same methods as for antisense delivery.

The antagonist or regulating agent or active ingredient is dosed and delivered in a pharmaceutically acceptable carrier as described herein below. The term antagonist or antagonizing is used in its broadest sense. Antagonism can include any mechanism or treatment which results in inhibition, inactivation, blocking or reduction in gene activity or gene product and for example preventing progression from non-invasive to invasive. The antagonizing step can include blocking cellular receptors for the gene products and can include antisense treatment as discussed herein.

Many reviews have covered the main aspects of antisense (AS) technology and its enormous therapeutic potential (Wright and Anazodo, 1995).

There are reviews on the chemical (Crooke, 1995; Uhlmann et al, 1990), cellular (Wagner, 1994) and therapeutic (Hanania, et al, 1995; Scanlon, et al, 1995; Gewirtz, 1993) aspects of this rapidly developing technology. Antisense intervention in the expression of specific genes can be achieved by the use of synthetic AS oligonucleotide sequences (for recent reports see Lefebvre-d'Hellencourt et al, 1995; Agrawal, 1996; Lev-Lehman et al, 1997). AS oligonucleotide sequences can be short sequences of DNA, typically 15-30 mer but can be as small as 7 mer (Wagner et al, 1996), designed to complement a target mRNA of interest and form an RNA:AS duplex. (See also Calabretta et al, 1996). Phosphorothioate antisense oligonucleotides do not normally show significant toxicity at concentrations that are effective, exhibit sufficient pharmacodynamic half-lives in animals (Agarwal et al., 1996) and are nuclease resistant. Instead of an antisense sequences as discussed herein above, ribozymes can be utilized. This is particularly necessary in cases where antisense therapy is limited by stoichiometric considerations (Sarver et al., 1990, Gene Regulation and Aids, pp. 305-325). (See also Hampel and Tritz, 1989; Uhlenbeck, 1987).

Ribozymes catalyze the phosphodiester bond cleavage of RNA. Several ribozyme structural families have been identified including Group I introns, RNase P, the hepatitis delta virus ribozyme, hammerhead ribozymes and the hairpin ribozyme (Sullivan, 1994; U.S. Patent No. 5,225,347, columns 4-5). Modifications or analogues of nucleotides can be introduced to improve the

therapeutic properties of the nucleotides. Improved properties include increased

nuclease resistance and/or increased ability to permeate cell membranes.

Nuclease resistance, where needed, is provided by any method known in the art

that does not interfere with biological activity of the antisense oligodeoxy-

nucleotides, cDNA and/or ribozymes as needed for the method of use and

delivery (Iyer et al., 1990; Eckstein, 1985; Spitzer and Eckstein, 1988; Woolf et

al., 1990; Shaw et al., 1991). Modifications that can be made to oligonucleotides

in order to enhance nuclease resistance include, but are not limited to, modifying

the phophorous or oxygen heteroatom in the phosphate backbone. These

modifications also include preparing methyl phosphonates, phosphorothioates,

phosphorodithioates and morpholino oligomers.

The present invention also includes all analogues of, or

modifications to, an oligonucleotide or polynucleotide of the invention that does

not substantially affect the function of the oligonucleotide. The nucleotides can

be selected from naturally occurring or synthetic modified bases. Naturally

occurring bases include adenine, guanine, cytosine, thymine and uracil. Modified

bases of the oligonucleotides include xanthine, hypoxanthine, 2-aminoadenine,

6-methyl, 2-propyl and other alkyl adenines, 5-halo uracil, 5-halo cytosine, 6-aza

cytosine and 6-aza thymine, psuedo uracil, 4-thiuracil, 8-halo adenine, 8-

aminoadenine, 8-thiol adenine, 8-thiolalkyl adenines, 8-hydroxyl adenine and

other 8-substituted adenines, 8-halo guanines, 8-amino guanine, 8-thiol guanine,

8-thioalkyl guanines, 8-hydroxyl guanine and other substituted guanines, other

aza and deaza adenines, other aza and deaza guanines, 5-trifluoromethyl uracil and 5-trifluoro cytosine.

In addition, analogues of nucleotides and/or polynucleotides can be prepared wherein the structure of the nucleotide and/or polynucleotide is fundamentally altered and that are better suited as therapeutic or experimental reagents. An example of a nucleotide analogue is a peptide nucleic acid (PNA) wherein the deoxyribose (or ribose) phosphate backbone in DNA (or RNA) is replaced with a polyamide backbone which is similar to that found in peptides. PNA analogues have been shown to be resistant to degradation by enzymes and to have extended lives *in vivo* and *in vitro*. Further, PNAs have been shown to bind stronger to a complementary DNA sequence than a DNA molecule. This observation is attributed to the lack of charge repulsion between the PNA strand and the DNA strand. Other modifications that can be made to oligonucleotides include polymer backbones, cyclic backbones, or acyclic backbones.

The active ingredients of pharmaceutical compositions can include oligonucleotides that are nuclease resistant as are needed for the practice of the invention or a fragment thereof shown to have the same effect when targeted against the appropriate sequence(s) and/or ribozymes. Combinations of active ingredients as disclosed in the present invention can be used, including combinations of antisense sequences.

The antisense oligonucleotides (and/or ribozymes) and cDNA of the present invention can be synthesized by any method known in the art for ribonucleic or deoxyribonucleic nucleotides. For example, an Applied Biosystems

380B DNA synthesizer can be used. When fragments are used, two or more such sequences can be synthesized and linked together for use in the present invention.

The nucleotide sequences of the present invention can be delivered either directly or with viral or non-viral vectors. When delivered directly the sequences are generally rendered nuclease resistant. Alternatively the sequences can be incorporated into expression cassettes or constructs such that the sequence is expressed in the cell as discussed herein below. Generally the construct contains the proper regulatory sequence or promotor to allow the sequence to be expressed in the targeted cell.

The proteins of the present invention can be produced recombinantly (see generally Marshak et al, 1996 "Strategies for Protein Purification and Characterization. A laboratory course manual.", CSHL Press) and analogues can be due to post-translational processing.

More in particular, with respect to polynucleotides disclosed herein, and corresponding polypeptides expressed from them, the invention further comprehends isolated and/or purified polynucleotides (nucleic acid molecules) and isolated and/or purified polypeptides having at least about 70%, preferably at least about 75% homology , more preferably at least about 80% , even more preferably at least about 90% , most preferably at least about 95% homology to the polynucleotides and polypeptides disclosed herein.

Nucleotide sequence homology can be determined using the "Align" program of Myers and Miller, ((1988) CABIOS 4:11-17) and available at

19

NCBI. Alternatively or additionally, the term "homology" ," for instance, with respect to a nucleotide or amino acid sequence, can indicate a quantitative measure of homology between two sequences. The percent sequence homology can be calculated as (Nref - Ndif)*100/Nref , wherein Ndif is the total number of non-identical residues in the two sequences when aligned and wherein Nref is the number of residues in one of the sequences Hence, the DNA sequence AGTCAGTC has a sequence similarity of 75% to AATCAATC (Nref = 8; Ndif = 2).

Alternatively or additionally, "homology" with respect to sequences can refer to the number of positions with identical nucleotides or amino acid residues divided by the number of nucleotides or amino acid residues in the shorter of the two sequences wherein alignment of the two sequences can be determined in accordance with the Wilbur and Lipman algorithm ((1983) Proc. Natl. Acad. Sci. USA 80:726), for instance, using a window size of 20 nucleotides, a word length of 4 nucleotides, and a gap penalty of 4, and computer-assisted analysis and interpretation of the sequence data including alignment can be conveniently performed using commercially available programs (e.g., Intelligenetics™ Suite, Intelligenetics Inc., CA). When RNA sequences are said to be similar, or to have a degree of sequence identity or homology with DNA sequences, thymidine (T) in the DNA sequence is considered equal to uracil (U) in the RNA sequence . RNA sequences within the scope of the invention can be derived from DNA sequences or their complements, by substituting thymidine (T) in the DNA sequence with uracil (U).

Additionally or alternatively, amino acid sequence similarity or identity or homology can be determined, for instance, using the BlastP program (Altschul et al. Nucl. Acids Res. 25:3389-3402) and available at NCBI. The following references provide algorithms for comparing the relative percentage homology of amino acid residues of two proteins, and additionally, or alternatively, with respect to the foregoing, the teachings in these references can be used for determining percent homology: Smith et al. (1981) Adv. Appl. Math. 2:482-489; Smith et al. (1983) Nucl. Acids Res. 11:2205-2220; Devereux et al. (1984) Nucl. Acids Res. 12:387-395; Feng et al. (1987) J. Molec. Evol. 25:351-360; Higgins et al. (1989) CABIOS 5:151-153; and Thompson et al. (1994) Nucl. Acids Res. 22:4673-480.

Polynucleotide sequences that are complementary to any of the sequences or fragments encompassed by the present invention discussed above are also considered to be part of the present invention. Whenever any of the sequences discussed above are produced in a cell, the complementary sequence is concomitantly produced and, thus, the complementary sequence can also be used as a probe for the same diagnostic purposes.

"Functionally relevant" refers to the biological property of the molecule and in this context means an *in vivo* effector or antigenic function or activity that is directly or indirectly performed by a naturally occurring protein or nucleic acid molecule. Effector functions include, but are not limited to include, receptor binding, any enzymatic activity or enzyme modulatory activity, any carrier binding activity, any hormonal activity, any activity in promoting or inhibiting adhesion of cells to

extracellular matrix or cell surface molecules, or any structural role as well as having the nucleic acid sequence encode functional protein and can be expressible. The antigenic functions essentially mean the possession of an epitope or antigenic site that is capable of cross-reacting with antibodies raised against a naturally occurring protein. Biologically active analogues share an effector function of the native which can, but do not necessarily, additionally possess an antigenic function.

The above discussion provides a factual basis for the use of the sequences of the present invention to identify bladder cancer-associated genes and provide diagnostic probes and markers to identify bladder cancer, particularly in the early stages of TCC.

EXAMPLES

EXAMPLE 1

METHODS OF THE INVENTION

A detailed description of the methods employed in the present invention is set forth in co-assigned US patent application USSN 09/534,661 filed on March 24, 2000, corresponding to PCT patent publication number WO 00/56935 and incorporated herein by reference in its entirety. The method includes preparing cell fractionations; extracting intact total RNA from membrane bound polysomes and free polysomes; preparing cDNA probes from template RNA derived from the extracted polysomes; performing microarray-based

22

comparison of the relative abundance of the different RNA species; analyzing the results; thereby identifying genes or clones encoding specifically membranal or secreted proteins.

Identification of cDNAs and genes encoding secreted or membranal encoding mRNAs is of major importance in TCC. More specifically, novel genes which mark the early stages of TCC and code for secreted proteins are the ultimate markers for diagnosis and follow-up of TCC. By deriving probes from template RNA extracted from membrane-bound polysomes and free polysomes and performing microarray-based comparison of the relative abundance of different RNA species, such potentially secreted proteins can be identified. Analysis of the results of such comparison and identification of the clones encoding for membranal or secreted proteins provides a valuable tool which can be used together with other gene discovery tools, and which in itself enables identification of likely targets for drug development.

Since membranal and secreted proteins are both accessible and critical for transduction of numerous intra- and intercellular signals, they are generally viewed as preferred targets for pharmacological use and intervention. Therefore, the a priori classification of arrayed unknown gene sequences into those that potentially code for secreted and membranal proteins is of great value for the optimization of a high-throughput process of identifying potential drug targets. Furthermore, the identification of genes which express membranal or secreted proteins that are differentially expressed in different cellular situations is of the utmost importance in designing therapeutic or diagnostic tools for TCC.

A method of identifying clones which encode membranal and secreted proteins was employed by preparing bladder cancer cell fractionations, preparing cDNA probes from template RNA derived from membrane-bound polysomes and free-polysomes, performing a microarray-based comparison of the relative abundance of different RNA species, analyzing the results and thereby identifying genes encoding for membranal and secreted proteins. Since membranal and secreted proteins are generally viewed as preferred targets for pharmacological intervention, the present invention thus provides a method of identifying likely targets for TCC diagnosis and therapy.

HYBRIDIZATION AND PROBES:

TCC and normal bladder hybridization:

The probes were prepared from normal healthy bladder samples and from TCC tumors. Only intact RNA with a proper histological report indicating the existence of TCC was used. All normal and tumor material was collected from two separate clinical centers. Such approach minimizes the influence of local specific surgical bias or subjectivity of the pathological report. Forty-one hybridizations were performed. In each hybridization, two probes were used simultaneously, each labeled with either Cy3 or Cy5.

These probes were as follows: Probe 1. Probe 1 was common to all hybridizations (common control probe). RNA from TCC samples was mixed with RNA from normal bladder samples. An equal amount of the RNA mixture was labeled with Cy3 and used in all hybridizations; and Probe 2. In each of the

hybridizations, a different RNA sample from a single donor was used (test probe).

A common control for all the hybridizations enables comparison of the results between the different hybridizations. If the common control (probe 1) hybridization results are similar in pattern in different hybridizations, comparison can be made between the results of probe 2 hybridizations and all hybridizations. Seventeen hybridizations included 16 RNA samples extracted from different control healthy bladder mucosa labeled with Cy5. Twenty-three hybridizations were performed with RNA samples derived from tumor tissues, either from non-invasive Ta or from T1 stages of development. Two hybridizations were performed with RNA extracted from 2 invasive TCC samples.

The hybridizations were carried out in three separate sets, but the same common control was used in all sets. Set 1 includes hybridizations 2-11 (TC2-TC11), set 2 includes hybridizations 16-25 (TC16-TC25), and set 3 includes hybridizations 28-41 (TC28-TC41). By using three different sets of hybridizations, the possibility of technical effects related to specific hybridizations is reduced. See Tables below and related description.

Probe from annotation of potentially secreted proteins:

TCC cell line -T24- (from ATCC) was used for cellular fractionation. Membrane-bound polysomes were separated from free polysomes using a sucrose step gradient. RNA coding for potentially secreted proteins was isolated

from this microsomal-membranal fraction and separated from RNA coding for intracellular proteins. Hybridization was performed as described hereunder. The probes used were as follows: Probe 1. Free polysomal RNA fraction labeled with Cy3; and Probe 2. Membrane-bound RNA fraction with Cy5.

TCC CHIP PREPARATION

All hybridizations were performed on TCC designated microarray. The microarray was made up of cDNA clones derived from 3 different libraries: SDGI library: (Described in co-assigned US Patent Application USSN 09/538,709, filed 30 March, 2000, corresponding to PCT application filed March, 2001 and incorporated herein by reference in its entirety): A pool of non-invasive TCC, invasive TCC and normal bladder was used for library preparation. 4550 clones from the SDGI library were included in the TCC chip.

Antisense library: (Described in co-assigned US Provisional Patent Application SN 60/157,843 , filed 6 October, 1999, corresponding to PCT application PCT/US00/27557, filed 6 October, 2000, and incorporated herein by reference in its entirety): The same cDNA pool used for the SDGI library was used for the preparation of a library enriched for antisense sequences. 450 clones from this library were included in the TCC chip.

SSH library: (Diatchenko et al., 1996). A subtraction library was made as follows. A normal bladder RNA pool was used for subtraction from non-invasive TCC RNA pool. The subtracted cDNA was used for the microarray printing. 5000 clones from the SSH library were used for printing.

<u>General methods in molecular biology:</u>

Standard molecular biology techniques known in the art and not specifically described were generally followed as in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York (1989), and in Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley and Sons, Baltimore, Maryland (1989) and in Perbal, *A Practical Guide to Molecular Cloning*, John Wiley & Sons, New York (1988), and in Watson et al., *Recombinant DNA*, Scientific American Books, New York and in Birren et al (eds) *Genome Analysis: A Laboratory Manual Series, Vols. 1-4* Cold Spring Harbor Laboratory Press, New York (1998) and methodology as set forth in United States patents 4,666,828; 4,683,202; 4,801,531; 5,192,659 and 5,272,057 and incorporated herein by reference. Polymerase chain reaction (PCR) was carried out generally as in *PCR Protocols: A Guide To Methods And Applications*, Academic Press, San Diego, CA (1990). In-situ (In-cell) PCR in combination with Flow Cytometry can be used for detection of cells containing specific DNA and mRNA sequences (Testoni et al, 1996, Blood 87:3822.) General methods in immunology: Standard methods in immunology known in the art and not specifically described are generally followed as in Stites et al.(eds), Basic and Clinical Immunology (8th Edition), Appleton & Lange, Norwalk, CT (1994) and Mishell and Shiigi (eds), Selected Methods in Cellular Immunology, W.H. Freeman and Co., New York (1980).

Immunoassays

In general, ELISAs where appropriate are one of the immunoassays employed to assess a specimen. ELISA assays are well known to those skilled in the art. Both polyclonal and monoclonal antibodies can be used in the assays. Where appropriate other immunoassays, such as radioimmunoassays (RIA) can be used as are known to those in the art. Available immunoassays are extensively described in the patent and scientific literature. See, for example, United States patents 3,791,932; 3,839,153; 3,850,752; 3,850,578; 3,853,987; 3,867,517; 3,879,262; 3,901,654; 3,935,074; 3,984,533; 3,996,345; 4,034,074; 4,098,876; 4,879,219; 5,011,771 and 5,281,521 as well as Sambrook et al, *Molecular Cloning: A Laboratory Manual*, Cold Springs Harbor, New York, 1989

Antibody Production

Antibodies can be either monoclonal, polyclonal or recombinant. Conveniently, the antibodies can be prepared against the immunogen or portion thereof for example a synthetic peptide based on the sequence, or prepared recombinantly by cloning techniques or the natural gene product and/or portions thereof can be isolated and used as the immunogen. Immunogens can be used to produce antibodies by standard antibody production technology well known to those skilled in the art as described generally in Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1988 and Borrebaeck, *Antibody Engineering - A Practical Guide*, W.H. Freeman

28

and Co., 1992. Antibody fragments can also be prepared from the antibodies and include Fab, F(ab')2, and Fv by methods known to those skilled in the art. For producing polyclonal antibodies a host, such as a rabbit or goat, is immunized with the immunogen or immunogen fragment, generally with an adjuvant and, if necessary, coupled to a carrier; antibodies to the immunogen are collected from the sera. Further, the polyclonal antibody can be absorbed such that it is monospecific. That is, the sera can be absorbed against related immunogens so that no cross-reactive antibodies remain in the sera rendering it monospecific.

For producing monoclonal antibodies the technique involves hyperimmunization of an appropriate donor with the immunogen, generally a mouse, and isolation of splenic antibody producing cells. These cells are fused to a cell having immortality, such as a myeloma cell, to provide a fused cell hybrid which has immortality and secretes the required antibody. The cells are then cultured, in bulk, and the monoclonal antibodies harvested from the culture media for use.

For producing recombinant antibody (see generally Huston et al, 1991; Johnson and Bird, 1991; Mernaugh and Mernaugh, 1995), messenger RNAs from antibody producing B-lymphocytes of animals, or hybridoma are reverse-transcribed to obtain complementary DNAs (CDNAs). Antibody cDNA, which can be full or partial length, is amplified and cloned into a phage or a plasmid. The cDNA can be a partial length of heavy and light chain cDNA, separated or connected by a linker, e.g., encoding a single chain antibody. The

antibody, or antibody fragment, is expressed using a suitable expression system to obtain recombinant antibody. Antibody cDNA can also be obtained by screening pertinent expression libraries.

The antibody can be bound to a solid support substrate or conjugated with a detectable moiety or be both bound and conjugated as is well known in the art. (For a general discussion of conjugation of fluorescent or enzymatic moieties see Johnstone & Thorpe, *Immunochemistry in Practice*, Blackwell Scientific Publications, Oxford, 1982.) The binding of antibodies to a solid support substrate is also well known in the art. (see for a general discussion Harlow & Lane *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Publications, New York, 1988 and Borrebaeck, *Antibody Engineering - A Practical Guide*, W.H. Freeman and Co., 1992) The detectable moieties contemplated with the present invention can include, but are not limited to, fluorescent, metallic, enzymatic and radioactive markers such as biotin, gold, ferritin, alkaline phosphatase, β-galactosidase, peroxidase, urease, fluorescein, rhodamine, tritium, ^{14}C and iodination.

Recombinant Protein Purification

Marshak et al, "Strategies for Protein Purification and Characterization. A laboratory course manual." CSHL Press, 1996.

Gene therapy:

The genes described in this patent can also be used as targets for gene therapy, since these genes can be of importance for the development of TCC. Therefore, targeted gene therapy against one or more of these genes , or against one or more of the corresponding polypeptides encoded by these genes, is applied to cure TCC and /or to retard the spread of TCC. BGene therapy as used herein refers to the transfer of genetic material (e.g. DNA or RNA) of interest into a host to treat or prevent a genetic or acquired disease or condition phenotype. The genetic material of interest encodes a product (e.g. a protein, polypeptide, peptide, functional RNA, antisense) whose production *in vivo* is desired. For example, the genetic material of interest can encode a hormone, receptor, enzyme, polypeptide or peptide of therapeutic value. Alternatively, the genetic material of interest encodes a suicide gene. For a review see, in general, the text "Gene Therapy" (Advances in Pharmacology 40, Academic Press, 1997).

Vectors can be introduced into cells or tissues by any one of a variety of known methods within the art. Such methods can be found generally described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Springs Harbor Laboratory, New York (1989, 1992), in Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley and Sons, Baltimore, Maryland (1989), Chang et al., *Somatic Gene Therapy*, CRC Press, Ann Arbor, MI (1995), Vega et al., *Gene Targeting*, CRC Press, Ann Arbor, MI (1995), *Vectors: A Survey of Molecular Cloning Vectors and Their Uses*, Butterworths, Boston MA (1988) and Gilboa et al (1986) and include, for example, stable or transient

31

transfection, lipofection, electroporation and infection with recombinant viral vectors. In addition, see United States patent 4,866,042 for vectors involving the central nervous system and also United States patents 5,464,764 and 5,487,992 for positive-negative selection methods.

Introduction of nucleic acids by infection offers several advantages over the other listed methods. Higher efficiency can be obtained due to their infectious nature. Moreover, viruses are very specialized and typically infect and propagate in specific cell types. Thus, their natural specificity can be used to target the vectors to specific cell types *in vivo* or within a tissue or mixed culture of cells. Viral vectors can also be modified with specific receptors or ligands to alter target specificity through receptor mediated events.

A specific example of DNA viral vector for introducing and expressing recombinant sequences is the adenovirus derived vector Adenop53TK. This vector expresses a herpes virus thymidine kinase (TK) gene for either positive or negative selection and an expression cassette for desired recombinant sequences. This vector can be used to infect cells that have an adenovirus receptor which includes most cancers of epithelial origin as well as others. This vector as well as others that exhibit similar desired functions can be used to treat a mixed population of cells and can include, for example, an *in vitro* or *ex vivo* culture of cells, a tissue or a human subject.

Additional features can be added to the vector to ensure its safety and/or enhance its therapeutic efficacy. Such features include, for example, markers that can be used to negatively select against cells infected with the

recombinant virus. An example of such a negative selection marker is the TK gene described above that confers sensitivity to the antibiotic gancyclovir. Negative selection is therefore a means by which infection can be controlled because it provides inducible suicide through the addition of antibiotic. Such protection ensures that if, for example, mutations arise that produce altered forms of the viral vector or recombinant sequence, cellular transformation can not occur.

Features that limit expression to particular cell types can also be included. Such features include, for example, promoter and regulatory elements that are specific for the desired cell type.

In addition, recombinant viral vectors are useful for *in vivo* expression of a desired nucleic acid because they offer advantages such as lateral infection and targeting specificity. Lateral infection is inherent in the life cycle of, for example, retrovirus and is the process by which a single infected cell produces many progeny virions that bud off and infect neighboring cells. The result is that a large area becomes rapidly infected, most of which was not initially infected by the original viral particles. This is in contrast to vertical-type of infection in which the infectious agent spreads only through daughter progeny. Viral vectors can also be produced that are unable to spread laterally. This characteristic can be useful if the desired purpose is to introduce a specified gene into only a localized number of targeted cells.

As described above, viruses are very specialized infectious agents that have evolved, in many cases, to elude host defense mechanisms. Typically,

viruses infect and propagate in specific cell types. The targeting specificity of viral vectors utilizes its natural specificity to specifically target predetermined cell types and thereby introduce a recombinant gene into the infected cell. The vector to be used in the methods of the invention depends on desired cell type to be targeted and is known to those skilled in the art. Thus, if bladder cancer is to be treated then a vector specific for such epithelial cells are used.

Retroviral vectors can be constructed to function either as infectious particles or to undergo only a single initial round of infection. In the former case, the genome of the virus is modified so that it maintains all the necessary genes, regulatory sequences and packaging signals to synthesize new viral proteins and RNA. Once these molecules are synthesized, the host cell packages the RNA into new viral particles which are capable of undergoing further rounds of infection. The vector's genome is also engineered to encode and express the desired recombinant gene. In the case of non-infectious viral vectors, the vector genome is usually mutated to destroy the viral packaging signal that is required to encapsulate the RNA into viral particles. Without such a signal, any particles that are formed do not contain a genome and therefore cannot proceed through subsequent rounds of infection. The specific type of vector depends upon the intended application. The actual vectors are also known and readily available within the art or can be constructed by one skilled in the art using well-known methodology.

The recombinant vector can be administered in several ways. If viral vectors are used, for example, the procedure can take advantage of their

target specificity and consequently, do not have to be administered locally at the diseased site. However, local administration can provide a quicker and more effective treatment, administration can also be performed by, for example, intravenous or subcutaneous injection into the subject. Following injection, the viral vectors circulate until they recognize host cells with the appropriate target specificity for infection.

An alternate mode of administration can be by direct inoculation into the bladder i.e., locally to the site of the disease or pathological condition or by inoculation into the vascular system supplying the site with nutrients. Local administration is advantageous because there is no dilution effect and, therefore, a smaller dose is required to achieve expression in a majority of the targeted cells. Additionally, local inoculation can alleviate the targeting requirement required with other forms of administration since a vector can be used that infects all cells in the inoculated area. If expression is desired in only a specific subset of cells within the inoculated area, then promoter and regulatory elements that are specific for the desired subset can be used to accomplish this goal. Such non-targeting vectors can be, for example, viral vectors, viral genome, plasmids, phagemids and the like. Transfection vehicles such as liposomes can also be used to introduce the non-viral vectors described above into recipient cells within the inoculated area. Such transfection vehicles are known by one skilled within the art.

Chemical compounds

35

The chemical compounds to be administered comprise <u>inter alia</u> small chemical molecules; antibodies of all types or fragments thereof including single chain antibodies; antisense oligonucleotides, antisense oligonucleotides, polynucleotides, DNA or RNA molecules; proteins, polypeptides and peptides including peptido-mimetics and dominant negative peptides; ribozymes ; and expression vectors

Delivery of chemical compound

The compound of the present invention is administered and dosed in accordance with good medical practice, taking into account the clinical condition of the individual patient, the site and method of administration, scheduling of administration, patient age, sex, body weight and other factors known to medical practitioners. The pharmaceutically "effective amount" for purposes herein is thus determined by such considerations as are known in the art. The amount must be effective to achieve improvement including but not limited to improved survival rate or more rapid recovery, or improvement or elimination of symptoms and other indicators as are selected as appropriate measures by those skilled in the art.

In the method of the present invention, the compound of the present invention can be administered in various ways. It should be noted that it can be administered as the compound or as pharmaceutically acceptable salt and can be administered alone or as an active ingredient in combination with pharmaceutically acceptable carriers, diluents, adjuvants and vehicles. The compounds can be administered intravesically (directly into the bladder),orally,

subcutaneously or parenterally including intravenous, intraarterial, intramuscular, intraperitoneally, and intranasal administration as well as intrathecal and infusion techniques. Implants of the compounds are also useful. The patient being treated is a warm-blooded animal and, in particular, mammals including man. The pharmaceutically acceptable carriers, diluents, adjuvants and vehicles as well as implant carriers generally refer to inert, non-toxic solid or liquid fillers, diluents or encapsulating material not reacting with the active ingredients of the invention.

It is noted that humans are treated generally longer than the mice or other experimental animals exemplified herein which treatment has a length proportional to the length of the disease process and drug effectiveness. The doses can be single doses or multiple doses over a period of several days, but single doses are preferred.

The doses can be single doses or multiple doses over a period of several days. The treatment generally has a length proportional to the length of the disease process and drug effectiveness and the patient species being treated.

When administering the compound of the present invention parenterally, it is generally formulated in a unit dosage injectable form (solution, suspension, emulsion). The pharmaceutical formulations suitable for injection include sterile aqueous solutions or dispersions and sterile powders for reconstitution into sterile injectable solutions or dispersions. The carrier can be a solvent or dispersing medium containing, for example, water, ethanol, polyol (for

example, glycerol, propylene glycol, liquid polyethylene glycol, and the like),

suitable mixtures thereof, and vegetable oils.

Proper fluidity can be maintained, for example, by the use of a

coating such as lecithin, by the maintenance of the required particle size in the

case of dispersion and by the use of surfactants. Nonaqueous vehicles such a

cottonseed oil, sesame oil, olive oil, soybean oil, corn oil, sunflower oil, or peanut

oil and esters, such as isopropyl myristate, can also be used as solvent systems

for compound compositions. Additionally, various additives which enhance the

stability, sterility, and isotonicity of the compositions, including antimicrobial

preservatives, antioxidants, chelating agents, and buffers, can be added.

Prevention of the action of microorganisms can be ensured by various

antibacterial and antifungal agents, for example, parabens, chlorobutanol,

phenol, sorbic acid, and the like. In many cases, it is desirable to include isotonic

agents, for example, sugars, sodium chloride, and the like. Prolonged absorption

of the injectable pharmaceutical form can be brought about by the use of agents

delaying absorption, for example, aluminum monostearate and gelatin.

According to the present invention, however, any vehicle, diluent, or additive

used have to be compatible with the compounds.

Sterile injectable solutions can be prepared by incorporating the

compounds utilized in practicing the present invention in the required amount of

the appropriate solvent with various of the other ingredients, as desired.

A pharmacological formulation of the present invention can be administered to

the patient in an injectable formulation containing any compatible carrier, such as

various vehicle, adjuvants, additives, and diluents; or the compounds utilized in the present invention can be administered parenterally to the patient in the form of slow-release subcutaneous implants or targeted delivery systems such as monoclonal antibodies, vectored delivery, iontophoretic, polymer matrices, liposomes, and microspheres. Examples of delivery systems useful in the present invention include: 5,225,182; 5,169,383; 5,167,616; 4,959,217; 4,925,678; 4,487,603; 4,486,194; 4,447,233; 4,447,224; 4,439,196; and 4,475,196. Many other such implants, delivery systems, and modules are well known to those skilled in the art.

A pharmacological formulation of the compound utilized in the present invention can be administered orally to the patient. Conventional methods such as administering the compounds in tablets, suspensions, solutions, emulsions, capsules, powders, syrups and the like are usable. Known techniques which deliver it orally or intravenously or directly to the bladder (intravesically) and retain the biological activity are preferred. In one embodiment, the compound of the present invention can be administered initially by intravenous injection to bring blood levels to a suitable level. The patient's levels are then maintained by an oral dosage form, although other forms of administration, dependent upon the patient's condition and as indicated above, can be used. The quantity to be administered vary for the patient being treated and vary from about 100 ng/kg of body weight to 100 mg/kg of body weight per day and preferably are from 10 μg/kg to 10 mg/kg per day.

EXAMPLE 2

POLYNUCLEOTIDES AND DIAGNOSTIC APPLICATIONS

Utilizing the methods set forth above, the polynucleotides set forth in Tables I and 2 were identified and cloned as being differentially expressed in bladder cancer. 41 hybridizations were compared.

The polynucleotides described in Table I are identified by clone number and accession number. This list includes sequences of known genes whose function in bladder cancer was heretofore unknown and which were now found to upregulated in bladder cancer. Corresponding nucleic acid sequences are provided in Table 3.

The polynucleotides described in Table 2 are identified by clone number. This list includes sequences of novel genes which have no identity to known proteins or genes in the gene databases. Corresponding nucleic acid sequences are provided in Table 4.

In both Tables I and 2, the differential expression pattern of the different hybridization probes is provided. In both Table I and 2, the genes listed were found to be upregulated in at least 60% of TCC samples and unchanged in at least 75% of the normal samples.

Tables I and 2 show the genes as described in biological NCBI databases, with the Genebank number of each gene (where applicable) as presented in the NCBI database. The location of the clone in the TCC microarray of the present invention is set forth in the tables, with their clone ID in the TCC chip.

The expression differentials described in Tables I and 2 were calculated as follows: Since a common control probe was used for all hybridizations and the hybridizations were carried out in three separate sets, the expression differentials in each respective set were calculated as compared to one of the normal bladder samples, as a reference probe.

Thus, hybridization set 1 which includes hybridizations TC2-TC11, all the results are shown as compared to the TC7 (normal) hybridization result. In hybridization set 2 which includes hybridizations TC16-TC25, all the results were calculated in comparison to the TC22 (normal) hybridization result. In set 3, which includes hybridizations TC28-TC41, all the results were calculated compared to the reference normal probe from TC47.

EXAMPLE 3: KEY GENES

In the present invention, the results of the 41 hybridizations were analyzed on the TCC microarray, in order to provide a statistically meaningful set of genes (which each include one of the polynucleotides identified) that can identify TCC samples and be used as a TCC marker set. As a result, a sub-set of twenty-two (22) potential molecular markers for non-invasive TCC was identified and validated using supervised statistical analysis methods. The 22 genes identified as potential markers (listed in Table 5) code for secreted factors, cytoskeletal and membranal proteins, all potentially suitable for the development of non-invasive diagnostic tests. This marker set of genes is described below, in

41

Example 4, Section 5, entitled "Expression patterns, scores and significance values for 22 short-listed genes", and related Tables C1 and C2.

Thirteen (13) of these 22 polynucleotides are described in Example 2 (see Tables 1-4), and nine (9) are newly described in this Example (see Tables 5-6). In Table 5, the polynucleotides already described in Tables 1 and 2 are designated with an "x".

The 22 gene marker set was identified following reanalysis of the 41 hybridizations. All the experiments were constructed so that such an analysis can be performed. The hybridization scheme (described in Section 1 hereunder) was based on both individual sample hybridizations and a common control approach. All the hybridizations passed quality control examination and pre-processing steps (described in Sections 2 and 3 hereunder) which are critical to establish input material suitable for any statistical analysis. Following these pre-processing steps, the hybridization data was scored according to its "similarity" to the desired discrimination - non-invasive TCC *versus* normal urothelium (see Example 5).

Two independent (though related) standard scoring methods were used and individual genes were selected that discriminate between non-invasive TCC and normal urothelium (see Example 5, Section 4).

Full bioinformatic annotation analysis of the 22 listed genes (see Example 5, Section 6) and detailed description of their potential biological relevance to cancer in general and to TCC in particular is presented herein (see Example 5, Sections 6 and 7).

Based on the sequence annotation, 17 of the selected 22 markers were found to be known, characterized genes and 5 code for genes with an unknown protein product. At least four of the known genes and 1 of the genes with yet unknown protein products code for membranal or secreted proteins, based on the Applicants' proprietary "secreted" probe (described in co-assigned USSN 09/534, corresponding to PCT patent publication number WO 00/56935 which is incorporated herein by reference in its entirety) and on domain analysis. (See Example 5, Sections 4 and 6). Being secreted, some of these proteins can be identified in body fluids (in particular urine), thus alleviating the need for invasive tests. All other non-secreted proteins can also be detected in urine, which always contains shedded urothelial cells.

For a diagnostic assay, urine samples of TCC patients and of non-TCC patients should be analyzed. Urine can be collected, preserved in –70°C and used either for protein assays (Western analysis) with the relevant antibody and/or for ELISA tests with the same relevant antibody. Similarly, blood samples from the same donors can be collected, and the separated serum samples can be used for detection of the candidate proteins in the serum using similar protein analysis approach. After establishing the particular assay for a single protein, assays for a combination of 2 or more different proteins can be set to increase the validity of the obtained results for each sample and to obtain robustness.

According to biomedical literature, the 17 known genes were classified into three functional groups: tumorigenesis, keratinocyte differentiation, and cell motility and proliferation. Thus, these markers can also fulfill a functional

43

role in TCC. Being functionally relevant these genes can be used as possible targets for genes therapy for TCC. This can be achieved by antagonizing their affects in the tumor using antisense delivery approach (for all such proteins), blocking their enzymatic activity (for enzymes), or specific drug delivery, as relevant. Since different keratins were detected as being differentially expressed in TCC, specific typing of the different keratins in urine of TCC patients, using a single multi-gene assay, can facilitate and improve robust TCC diagnostics.

The specificity of markers for TCC over other cancers is an important consideration. In particular, the expression level of the selected gene set is analyzed in other urogenital cancers, such as renal carcinoma and prostate cancer. Importantly, samples obtained from clinically relevant controls, such as inflammation or benign prostate hyperplasia (BPH) must be included. Retrospective studies of patients is also be carried out, as well as comparison of samples obtained during follow-up procedures (to monitor tumor progression in TCC patients).

All the genes described in the present invention are tested for their level of expression of exfoliated cells in urine, according to the following protocol. Urine samples (e.g. 100ml of urine) are collected from 3 different populations: healthy donors, TCC patients and a relevant control group (e.g., prostate cancer, bladder inflammation). The exfoliated cells are separated from urine (it is possible to keep the separated cells in –70°C pending further work) and used for preparation of RNA. Such an approach enables tracking cancer-related changes at the level of gene expression. Following RNA extraction, RT-PCR is performed

for selected genes. Primers specific for each of these genes are constructed and used for the amplification of the cDNA products, being constructed so that each of the tested genes are be amplified to a fragment of a different size. RT-PCR reaction is carried out in semi-quantitative approach. Fractionation for the resultant products on gels indicates the relative abundance of each of the tested genes in the tested RNA sample. Alternatively, TaqMan (Applied Biosystems) enables a fully quantitative time-course demonstration of the level of expression of these genes. Results for each of the tested genes are defined and documented and used for statistical analysis. Finally, a value is calculated for the expression of the predictor gene set in TCC and in non-TCC samples. Comparison of the expression level results for a given unknown sample to this known calculated value predicts if the tested sample contains TCC (under certain confidence level ,p value). All information from the tested samples is gathered during the establishment of the described diagnostic protocol and the statistical analysis is expanded so that all samples participating in the study are included.

According to the present invention, gene sequences are included which are uniformly expressed in normal and TCC tissues (see Example 5, Section 8). These genes can be used as an internal control in each multiplex RT-PCR. To this end primers for amplification of such genes are constructed and applied within the RT-PCR reaction of the marker gene set.

In further analysis, the results obtained from a single donor are compared between different tissues obtained from the same donor, e.g., matched urine exfoliated cells and tumor tissue (for RT-PCR approach) or urine

and blood for the protein analysis approach. This enables a deeper understanding of the molecular changes associated with TCC and their general presentation in different organs.

The genes provided in the present patent application can also be used for printing a small diagnostic TCC mini-microarray. This chip includes also clones with a uniform high expression in both normal urothelium and TCC (see Example 5, Section 8). Such TCC mini-microarray can be used for both disease detection and validation and for molecular staging and grading of the TCC tumors. Samples for hybridization on such chip include material derived from TCC tumors and from normal urothelium from different donors.

In addition, urine exfoliated cells from the same donors can be used for RNA extraction and RNA amplification. The RNA can be used for generating cDNA probes for the TCC mini cDNA microarray. This enables characterization of gene expression patterns of all the printed genes and comparison between the expression pattern obtained using tissue-type material to that obtained by cells shedded in urine. Since shedded cells are also collected from non-TCC donors such as patients with BPH and inflammation, these also comprise part of the hybridization probes.

The present invention further presents the use of a small subset of genes (2 or more genes) together for providing an accurate diagnostic test for TCC. With one exception (the cytokeratin 8 and 18 assay), all commercially existing molecular diagnostics for TCC are based on tests for single proteins. These can be insufficient to account for the inherent complexity of cancer, as

well as for the variability of both healthy and affected populations. To this end, the present invention describes the use of a combination of several genes and/or proteins either as a marker set for detection of these proteins in urine or in other body fluids, and/or by using the cells or cell debris present in the urine of TCC patients for multiple-gene RT-PCR diagnostic testing.

In-situ hybridization analysis using the same gene set can be performed as an auxiliary qualitative validation step, using paraffin blocks from normal urothelium and TCC tumors.

The genes of the present invention also characterize different stages of TCC. Correct "staging" of TCC is fundamental for the management of this disease. Upon detection of a new TCC patient, the developmental stage of the tumors determines relevant treatment. For example, if a non-invasive tumor is identified, "TURT" is the surgical approach recommended. If, however, the tumor is defined as invasive TCC, cystectomy is usually the treatment of choice. Identification of those non-invasive TCC patients that might progress is of great clinical value.

Keratin 13 is identified herein as a marker that can differentiate Ta from T1 and invasive tumors (see Section 10). The analysis described in the present invention indicates a clear discrimination between Ta and T1 tumors, where this gene is upregulated in Ta tumors and downregulated in T1 tumors when compared to normal urothelium. This gene in include part of the diagnostic tests described.

47

According to the present invention, 22 polynucleotides included in 22 genes were identified; these genes serve as potential markers for TCC, especially for non-invasive TCC. These genes, and all the genes included in Tables 1 – 6, can be used for diagnosis of TCC.

Full-length genes or gene fragments are suggested as markers for non-invasive assays. PCR products, antisense products, protein products and antibodies raised against these genes can be applied both for diagnostics for TCC and for targeted gene therapy. The tests for the levels of these genes and/or proteins can be performed in body fluids, in the original tumor or in other relevant body organs, and in cells found in the urine of patients.

EXAMPLE 4 :HYBRIDIZATIONS AND STATISTICAL ANALYSIS

Section 1. Hybridization scheme

The hybridization scheme according to the present invention is based on three principles

1. Individual hybridization of each sample (normal or TCC) whenever possible: This provides a comprehensive overview of the entire sample set, with minimal a-priori assumptions, and with maximal measurement of the variability between the samples. Such individual hybridization procedure is crucial for successful analysis of the results. In a small number of cases, due to insufficient amounts of normal urothelium material, pools of several normal samples were used as a single probe (See Table A,3rd set).

2.   Utilization of an identical common normalizing probe

("Common Control" or "CC") in each set of hybridizations: By maintaining one of

the probes as a constant, common probe, hybridization results can be compared

across experiments. The common normalizing probe used in the present

invention was prepared from a pool of RNA from different TCC and normal

samples. This material should be similar in composition to the one used for

construction of the TCC microarray. Thus, it has a high probability to hybridize

and detect a maximal number of elements on the TCC array, and to provide an

appropriate normalization of signals between hybridizations.

3.   Secreted and membranal proteins have an obvious

advantage as molecular markers. The "secreted" probe allows sequence-

independent identification of genes potentially coding for secreted and

membranal proteins. If such genes are highly expressed in tumors it is plausible

to try to find their protein products highly expressed in urine, too.

Hybridizations of TCC and of normal urothelium samples were

performed in 3 sets which were separated in time as well as in the methods of

RNA preparation (polyA and total), as shown in Table A. Although these

differences increase the variability of the results, they also suggest that the

identified phenomena are robust to experimental intricacies. Comparison of gene

expression results between the sets increases the validity of the results obtained.

Differences in RNA preparation can also affect the common normalizing probe.

For example, in the first two sets an identical total RNA pool was used (Table A

common normalizing probe1), while in the 3rd set polyA RNA was extracted from

the same pool of total RNA and used as a common control (Table A, common

normalizing probe2).

Table A: Detailed hybridization scheme

Set number	Probe type	Probe 1 (Cy3)	Probe 2(Cy5): "Experiment" probe			
			Type (# samples)	Code	Stage	Grade
1	Total RNA	Common normalising probe1 (total RNA)	TCC (5)	TC2	T1	
				TC3	T1	G2/G3
				TC4	T1	G2/G3
				TC5	T1	
				TC6	T1	
			Normal (5)	TC7	normal	
				TC8	normal	
				TC9	normal	
				TC10	normal	
				TC11	normal	
2	Total RNA	Common normalising probe1 (total RNA)	TCC (6)	TC16	Ta	G2
				TC17	Ta	G2
				TC18	Ta	High
				TC19	T1	Low
				TC20	T1	G3
				TC25	T1	G1
			Normal (4)	TC22	normal	
				TC22	Normal	
				TC23	Normal	
				TC24	Normal	
3	Poly A RNA	Common normalising probe2 (poly A)	TCC (14)	TC28	T1+TIS	G3
				TC29	T1	High
				TC30	T1	G3
				TC31	Ta	G1/2
				TC32	Ta/T1	G2
				TC33	Ta	G2
				TC34	invasive	G3
				TC39	Ta	low
				TC40	T1	G1/2
				TC41	Ta	G2
				TC42	Ta	G2
				TC43	T1	G2
				TC44	Ta	G2
				TC45	invasive	G3
			Normal (19)	TC35	Normal	
				TC36	Normal	
				TC37	Normal pool	
				TC38	Normal pool	
				TC46	Normal	
				TC47	normal pool	
				TC48	normal pool	

Set number	Probe type	TCC invasive cell line	Code	Probe 1(Cy3)	Probe 2(Cy5)
4	Secreted	SW780	TC49	Free polysomal RNA	Membrane bound polysomal RNA
		T24	TC50		

50

In order to identify secreted and membranal proteins, two

"secreted" probes were prepared from human invasive TCC cell lines, T24 and

SW780. Briefly, membrane-bound polysomes were separated from free

polysomes using sucrose step gradient. RNA coding for potentially secreted

proteins was isolated from the microsomal–membranal fraction and RNA coding

for intracellular proteins from the free polysomal pellet. Each RNA ("Secreted"

and "Intracellular") was labelled with a different dye and hybridized to the TCC

array (Table A, set 4). Significant differential expression in one of the probes is

an indication of potential cellular compartments (intracellular or

secreted/membranal). As a convention, a negative differential represents

secreted proteins.

Section 2. Quality control (QC), preliminary evaluation

In order to ensure the quality of the results shows in the present

invention and to minimize experimental artefacts, all hybridization results

underwent several standard QC steps. Since the hybridizations were performed

in three separate sets, QC procedures were done *within* sets, consistent with

inventors' past experience.

1.   <u>Reproducibility of the common control probe</u>. Relative

expression levels are compared across hybridizations due to the use of a

common normalizing probe. However, this can be faithfully performed only if the

common control probe behaves consistently across each set of hybridization.

This consistency is first measured by the pair-wise correlations between the

51

common control signal vectors. The pair-wise correlation coefficient between common control probes in each set are almost invariably very high (>0.97, Table B1). These results indicate the suitability of common control-based normalization for this data set.

Table B1: <u>Common control correlations</u> (by hybridization set) for 41 TCC hybridizations

Set #1

	TC10	TC11	TC2	TC3	TC4	TC5	TC6	TC7	TC8	TC9
TC10	1.00	.96	.97	.97	.97	.98	.98	.97	.87	.98
TC11	.96	1.00	.97	.98	.97	.97	.98	.98	.88	.98
TC2	.97	.97	1.00	.98	.99	.98	.99	.98	.89	.98
TC3	.97	.98	.98	1.00	.99	.99	.98	.97	.88	.98
TC4	.97	.97	.99	.99	1.00	.99	.98	.97	.88	.98
TC5	.98	.97	.98	.99	.98	1.00	.98	.97	.89	.99
TC6	.98	.97	.98	.99	.97	.97	1.00	.98	.89	.98
TC7	.97	.97	.98	.98	.88	.88	.89	1.00	.89	.89
TC8	.87	.88	.88	.89	.88	.88	.99	.98	1.00	.89
TC9	.98	.98	.98	.98	.98	.98	.99	.98	.89	1.00

Set #2

	TC16	TC17	TC18	TC19	TC20	TC21	TC22	TC23	TC24	TC25
TC16	1.00	.98	.97	.98	.98	.97	.98	.97	.97	.97
TC17	.98	1.00	.98	.99	.98	.98	.98	.98	.96	.97
TC18	.97	.98	1.00	.98	.97	.97	.97	.97	.97	.97
TC19	.98	.99	.98	1.00	.98	.98	.98	.97	.97	.97
TC20	.98	.98	.97	.98	1.00	.98	.98	.97	.97	.98
TC21	.97	.98	.97	.98	.98	1.00	.98	.97	.97	.97
TC22	.98	.98	.97	.97	.97	.97	1.00	.98	.97	.97
TC23	.97	.98	.97	.97	.97	.97	.97	1.00	.97	.96
TC24	.97	.97	.96	.97	.97	.97	.98	.97	1.00	.96
TC25	.97	.98	.97	.97	.97	.97	.98	.97	.96	1.00

Set #3

	TC28	TC29	TC30	TC31	TC32	TC33	TC34	TC35	TC36	TC37	TC38	TC39	TC40	TC41	TC42	TC43	TC44	TC45	TC46	TC47	TC48
TC28	1.00	.99	.97	.98	.98	.98	.98	.98	.99	.99	.98	.97	.98	.97	.97	.98	.97	.97	.97	.98	.96
TC29	.99	1.00	.97	.97	.98	.98	.98	.98	.98	.98	.98	.98	.98	.97	.97	.98	.98	.97	.96	.97	.96
TC30	.97	.97	1.00	.98	.98	.98	.98	.98	.99	.99	.99	.98	.98	.98	.98	.97	.99	.98	.97	.98	.97
TC31	.98	.97	.98	1.00	.99	.98	.99	.98	.99	.99	.99	.98	.99	.98	.97	.98	.97	.96	.96	.97	.95
TC32	.98	.98	.98	.99	1.00	.99	.99	.98	.99	.99	.99	.98	.98	.98	.97	.97	.98	.97	.96	.97	.95
TC33	.98	.98	.98	.98	.99	1.00	.99	.98	.99	.99	.99	.98	.97	.97	.97	.96	.98	.96	.96	.96	.95
TC34	.98	.98	.98	.98	.99	.99	1.00	.98	.99	.99	.99	.98	.97	.97	.97	.96	.98	.96	.96	.98	.96
TC35	.98	.97	.98	.98	.98	.98	.98	1.00	.98	.99	.99	.98	.99	.98	.97	.99	.97	.97	.97	.98	.97
TC36	.99	.98	.99	.99	.99	.99	.99	.98	1.00	.99	.99	.99	.98	.98	.99	.97	.97	.97	.98	.97	.97
TC37	.99	.98	.98	.99	.99	.99	.99	.98	.99	1.00	.98	.99	.98	.98	.99	.97	.97	.98	.97	.98	.97
TC38	.99	.98	.98	.99	.99	.99	.99	.97	.98	.99	1.00	.99	.98	.98	.98	.98	.99	.98	.97	.98	.97
TC39	.98	.97	.98	.98	.98	.98	.98	.97	.99	.99	.99	1.00	.98	.98	.98	.97	.97	.97	.97	.97
TC40	.98	.98	.98	.98	.98	.99	.98	.98	.98	.98	.98	.98	1.00	.98	.98	.97	.98	.97	.97	.97
TC41	.97	.97	.97	.98	.98	.97	.97	.97	.96	.97	.98	.98	.98	.98	1.00	.97	.97	.97	.97	.97
TC42	.97	.97	.98	.98	.98	.99	.98	.98	.98	.99	.99	.99	.98	.98	1.00	.98	.97	.97	.98	.98
TC43	.98	.98	.98	.98	.99	.98	.98	.98	.99	.99	.99	.98	.98	.97	.97	1.00	.98	.97	.98	.99
TC44	.97	.97	.97	.97	.97	.98	.97	.97	.96	.97	.97	.97	.98	.98	.97	.98	1.00	.98	.98	.99
TC45	.97	.96	.97	.97	.97	.96	.96	.96	.97	.97	.97	.97	.97	.97	.97	.97	.98	1.00	.99	.99
TC46	.97	.97	.96	.97	.97	.96	.96	.95	.96	.97	.98	.98	.98	.98	.97	.97	.97	.98	.99	1.00	.99
TC47	.98	.97	.97	.97	.98	.97	.97	.96	.98	.98	.98	.97	.97	.97	.97	.97	.98	.99	.99	1.00
TC48	.96	.96	.96	.97	.97	.95	.95	.95	.96	.97	.97	.97	.97	.97	.97	.98	.99	.99	.99	1.00

2. <u>Signal quality</u>. A second measure of hybridization quality is the number of elements which yielded a significant signal and a reliable signal to background (S2B) ratio with each probe. Since a custom cDNA array was used, both experiment and control probes are expected to yield a similar number of significant signals (A set of n hybridizations of an m-gene array is typically treated as a matrix A of size mXn. Thus the expression level of each gene in all n hybridizations is a vector of size m ("gene vector")). A single hybridization experiment is represented by a vector of m expression measurements ("hybridization vector"). In differential profiling the vector can represent a single probe (in which case it is a vector of signals) or both (a vector of differentials)). Thus, when comparing the hybridization quality of the common control probes, pair-wise correlation was calculated between the common control signal vector in one hybridization to that of another. Missing values were deleted on a case-wise basis. Traditional threshold (200 units) for signals, for both common control and tested sample probes, and S2B (value of 2.5 for at least 40% coverage of element) were used (Table B2).

The first and third sets of hybridizations yielded signals of high quality in both common control and experiment probes. The quality of the second set was significantly lower (Table B2).

Table B2: Signal qua

Set #1

Code	Significant P1	Significant P2
TC10A	6679	6450
TC11A	5280	4546
TC2A	5834	5097
TC3A	6603	5691
TC4A	6319	4706
TC5A	6528	5589
TC6A	6773	6099
TC7A	6456	6016
TC8A	5762	5376
TC9A	6715	5929

TC46 A	6261	5030
TC47 A	6057	5206
TC48 A	5051	4109

Set #2

Code	Significant P1	Significant P2
TC16A	1762	1964
TC17A	1758	2063
TC18A	1543	1822
TC19A	1766	1987
TC20A	1578	1708
TC21A	1604	1540
TC22A	1700	1812
TC23A	1656	1832
TC24A	1417	1408
TC25A	1255	1555

Set #3

Code	Significant P1	Significant P2
TC28 A	8084	7690
TC29 A	7519	6887
TC30 A	7449	7404
TC31 A	7294	6697
TC32 A	7284	6529
TC33 A	7724	6919
TC34 A	7236	7282
TC35 A	7758	7258
TC36 A	7410	7089
TC37 A	7291	7583
TC38 A	7353	6760
TC39 A	6370	6321
TC40 A	6754	6722
TC41 A	6183	5502
TC42 A	5758	5501
TC43 A	7137	7013
TC44 A	5576	5335
TC45 A	4920	4917

3.   _Relationships between hybridizations_. Hierarchical clustering of the hybridizations in each set provides an additional, albeit preliminary, estimate of quality. Either the Pearson correlation coefficient or a standard Euclidean distance was used as the distance measure between differential hybridization vectors. Hybridizations were clustered according to these distances by average linkage hierarchical clustering. Missing values were deleted on a case-wise basis. Clusters of hybridizations can be identified and evaluated in light of existing knowledge.

Many of the unexpected phenomena can indicate the limitation of previous understanding, and serve as a starting point for class definition. However, "outlying" hybridizations can also indicate quality problems. Overall, in each set (Table B3) most of the separation between hybridizations is consistent with the expected TCC and normal urothelium separation. Even in hybridizations of lower quality, such as those of the second set, a clear separation between TCC and normal samples is observed.

One of the TCC samples in the first set (TC6) is such an "outlyer" (Table B3), as well as one of the normal samples in the second set, and another normal sample (TC35) in the third. (The "outlyers" do not appear to be misclassifications). For example, TC35 (a normal sample which is an "outlyer" in the third set) does not behave like a TCC sample. Rather those genes that are up-regulated in TCC samples are down-regulated in TC35).

SET #1

SET #2

Pearson correlation

Euclidean

56

Set#3

Tree Diagram for 21 Variables

Unweighted pair-group average
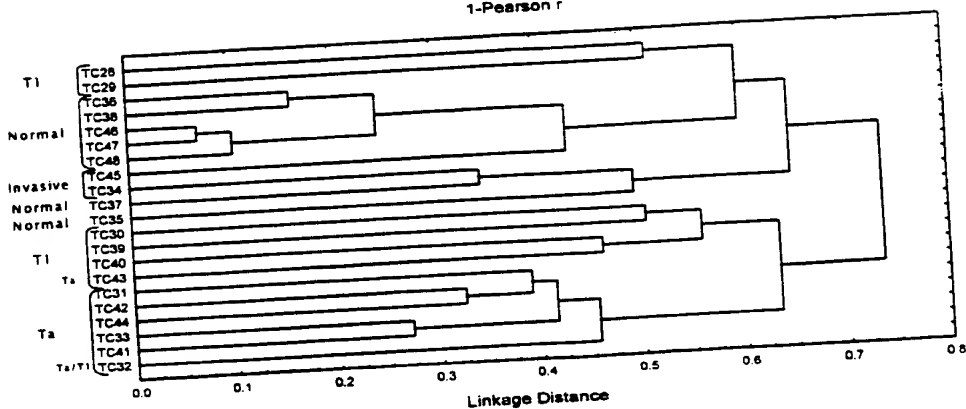
1-Pearson r

Pearson correlation

Tree Diagram for 21 Variables

Unweighted pair-group average

Euclidean distances

Euclidean

57

5

10

15

20

25

30

35

40

TCC All

Raw data

Unweighted pair-group average, 1-Pearson r

Normal

2nd set { Ta, T1

Normal

Invasive.Normal

Ta, T1

Ta

0.0   0.1   0.2   0.3   0.4   0.5   0.6   0.7   0.8

Linkage Distance

Pearson correlation

TCC All

Raw data

Unweighted pair-group average, Euclidean distances

Normal

Invasive.Normal

3rd set { T1 (Ta)

Ta

G3

10   20   30   40   50   60   70   80

Linkage Distance

Eucledean

45
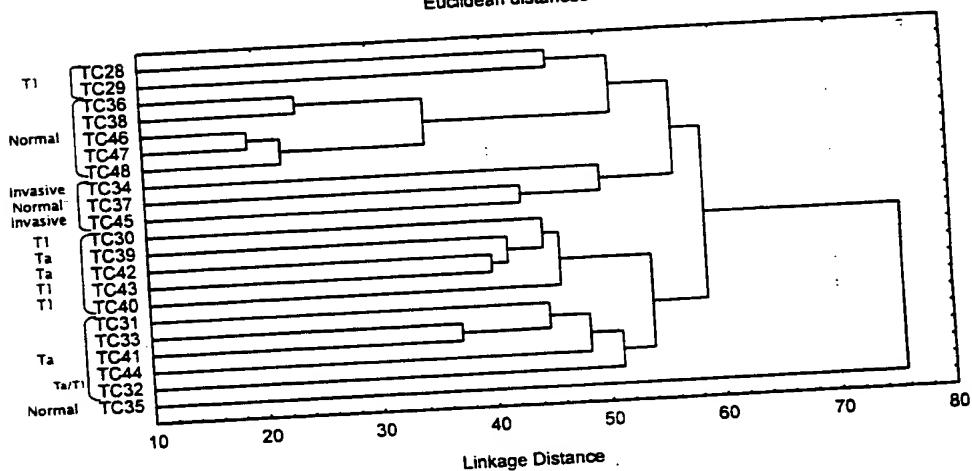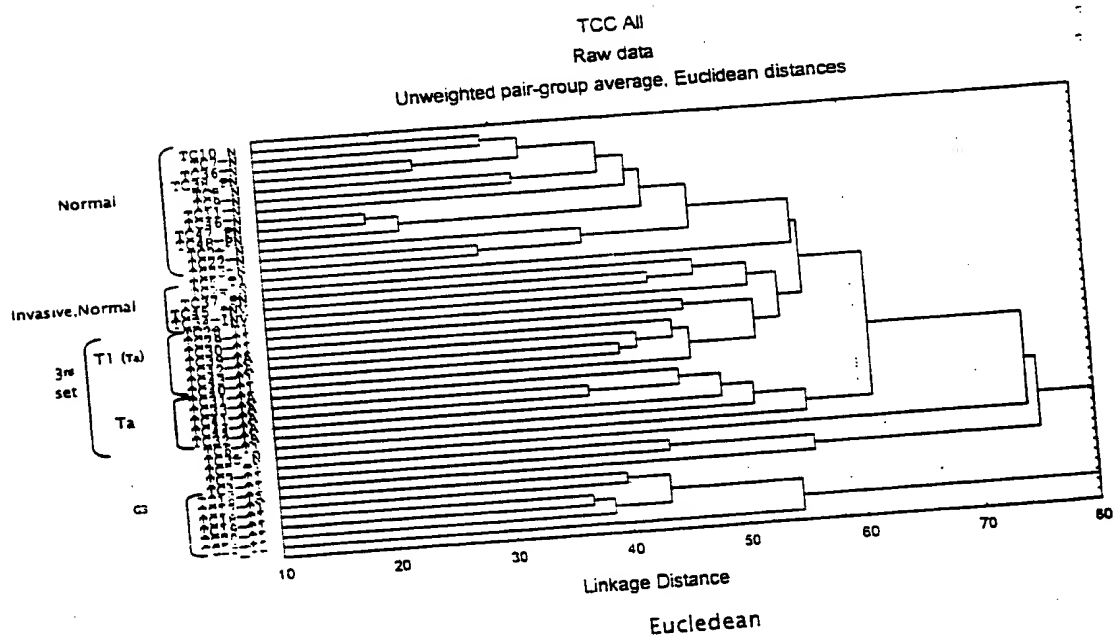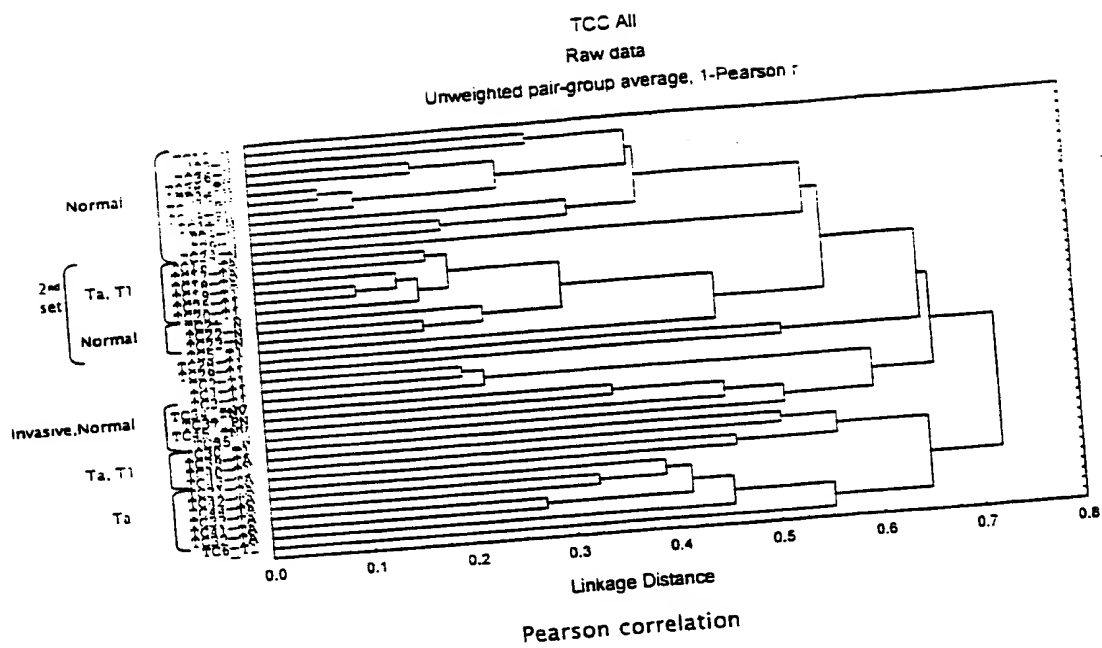
None of the "outlyers" was eliminated from subsequent analysis steps. Rather, they were included to facilitate the selection of a more robust marker set (Table B3). Here, more complex relations are observed between global expression profiles. First, the two invasive hybridizations (TC34 and TC45), are distinct from other TCC samples (Tables B3 and B4). Second, the relationship between global Ta and T1 profiles is not straightforward. Most of the Ta samples form a unique cluster in the 3rd set, while the T1 samples are more dispersed.

Section 3. Pre-processing of the hybridization data.

All hybridization data, even of good global quality, was filtered and processed prior to additional large scale analysis. This included global balancing of signals, identification and treatment of problematic signals, normalization of the hybridization data and filtering.

1. Signal Balancing. Differences in labeling and hybridization can bias the signals obtained with a Cy3 probe relative to the Cy5 probes. For each hybridization, linear balancing is used to overcome this bias. The balancing coeffient is calculated as (sum P1)/(Sum P2).

2. Problematic signals. Two types of problematic signals are identified: very low signals and exceptionally variable signals. The first are signals below a pre-set threshold. The second are common control signals which significantly (>2 SDs) deviate from the average common control signal for a given

59

element. All problematic common control signals were replaced with the average signals for the common control in the given element.

3. <u>Normalization</u>. In order to obtain meaningful differential expression values (TCC vs. normal) and to reduce differences between sets (inter-block variance), a second step of normalization is performed. In this step, each of the balanced differential expression levels (relative to the common control) is normalized by the average differential expression of the given element in the normal samples of the same set. The resulting normalized differential values give a measure of the difference between the expression of the element in a given sample (normal or TCC) and the average expression levels in normal tissues. Note that due to the use of averages in replacement of problematic signals, the variability in the normal samples is reduced by this procedure.

4. <u>Filtering.</u> In order to reduce the data set and limit it to higher-quality elements there was restricted from further analysis any overall weak elements (no signal above 200) and any non-differential elements (for which no normalized differential values exceeds |1.7|). The remaining number of elements following these filters is 6693. Low-quality elements (where more than 20% of signals are problematic) were filtered only in later stages.

Section 4. Class Prediction: Normal urothelium vs. non-invasive TCC and selection of marker set of genes

60

To discriminate between normal urothelium and non-invasive TCC, each of the gene with hybridization value is scored according to its "similarity" to the desired discrimination- N vs Non invasive TCC. Two independent (though related) standard scoring methods from the three described below were used.

Statistical Methods for Class prediction

Scoring methods

a.   *Student's unpaired t-test*: The t-test is a statistic for measuring the significance of a difference of the means between two distributions (m1, and m2) considering the variance (s21 and s22) within each group. The two populations are expected to be drawn from a normal distribution. In the case, these are the mean expression levels of a gene in normal urothelium and in TCC tumors, which are supposed to have a log-normal distribution (thus, log values are used). Statistical significance estimates (p-values) available for the t statistic. Since a large number of measurements is available for a small number of samples, a much more stringent threshold of significance is used ($p < 10^{-6}$), which, according to the Benferroni adjustment corresponds to $p < 0.001$.

a.   *estimation of prediction error*: This method scores genes according to probability of error or misclassification. As part of this procedure a discrimination threshold is determined. The threshold $T(g_j)$ is taken such that the two types of misclassification error become equal, as:

$$\text{(D.1)} \quad T(g_j) = [m1(g_j){*}s2(g_j) + m2(g_j){*}s1(g_j)] / [s1(g_j)+s2(g_j)]$$

61

and the significance of the misclassification error is given by

(D.4) $P = 1 - F[(T(g_j) - m2(g_j))/s2(g_j)]$

where F is a distribution function N(0,1).

b. _Receiver Operating Characteristics (ROC) curves._ ROC curves are used to evaluate the power of a classification method for different asymmetric weights of false negative vs. false positive errors (or sensitivity vs. specificity). In diagnostic applications false negative errors can be detrimental while false positives can be tolerated. A ROC curve plots the tradeoff between the two types of errors as the classification threshold varies. For each potential threshold, the rate of true positives is plotted against the rate of false positives. Accuracy (A) is indexed by the area under the curve. A straight line (i.e. 50:50 chance of correct diagnosis, no better than chance), has A=0.5. Perfect accuracy (A=1) means that for a given threshold all predictions are correct.

The first score used is the "student's unpaired t-test", as above-described, i.e., one-way ANOVA with two classes, which reflects the difference between the classes relative to the variance within classes. The distribution of this statistic is resolved and significance levels of each score (its p-value) can be derived. The second method used scored genes according to an "estimation of prediction error", as described above, which again provides significance estimates (p-values) in a straightforward way.

Both scoring methods yielded similar numbers of elements with statistically significant scores: 77 elements according to the t-test scores ($p < 10^{-6}$), and 63 elements with low misclassification errors.

This list was further narrowed according to several additional considerations:

1.   Exclusivity of up-regulated genes. Non-invasive tests, such as a urine test, require the identification of tumor cells or proteins on a considerable background of normal tissue. Since the inventors assume that only significantly up-regulated TCC genes have a chance to be detected on such background, while genes down-regulated in TCC are not be faithfully detected in non-invasive tests, they specifically selected such genes according to their normalized differentials. Furthermore, genes having a particularly low expression in normal tissues were prioritized, to minimize detection problems in further assays.

2.   Consistent scores. Elements with high scores in both methods were prioritized in the final list. The error-based method was given preference over the t-test scores due to the prediction thershold it provides.

3.   Redundancy. Approximately half of the clones on the array were derived by a subtraction procedure (SSH) enriching for TCC up-regulated clones. Inevitably, this significantly increases array redundancy, especially for up-regulated genes. In order to address this problem, a large portion of the significantly scored clones and up-regulated genes has been sequenced (~900

63

clones). Only a single representative of each redundancy group was retained in the list.

4. Element quality. A stricter threshold of element quality was added, and only elements with less than 8 problematic signals were included (in most genes a much smaller number of problematic signals was encountered).

5. Gene identity. The functional role fulfilled by different genes as well as previous knowledge can change their priority. For example, one low-scoring gene (FABP) was selected due to its involvement in psoriasis and squamous cell carcinoma of the bladder.

64

Section 5. Expression patterns, scores and significance values for the 22 short-listed genes of the invention

The final subset of informative genes comprises the top 22 up-regulated genes after application of consistency, redundancy, and quality filters. These genes obtained high scores as discriminators by both scoring methods.

The differential expression patterns, statistical scores and significance values for the 22 selected genes are shown in Table C1. The expression levels are shown in the Table in the following order:

Normal urothelium (16 first hybridizations);

T1 samples (13 hybridizations);

Ta samples (10 hybridizations).

The levels of differential expression and the signal values of the 2 "Secreted" probes are also shown (Table C1, four columns headed "Secreted..."). Strong negative differentials indicate a gene potentially encoding a secreted or membranal protein.

The statistical scores are given in the following order:

1. Estimation of mis-classification error is given under column "Error1_2", 2. P value - Fisher criteria (similar to T-test) is shown as P-values, (column "PvalueFisher"), and 3. ROC value in the column headed with the same name.

Table C2 includes the raw measured signals for each gene in 39 hybridizations. P1 are the common control signals. P2 signals represent the measured tissue samples, and are thus more interesting. The genes are sorted by statistical significance with the top gene having the highest score. This is also the order in which they have been incorporated into the predictor (Table D1).

Table D1: Unsupervised analysis of tumor samples

Hierarchical clustering of tumor samples (3rd set)

Classification of TCC tumors
5093 genes
Euclidean distances
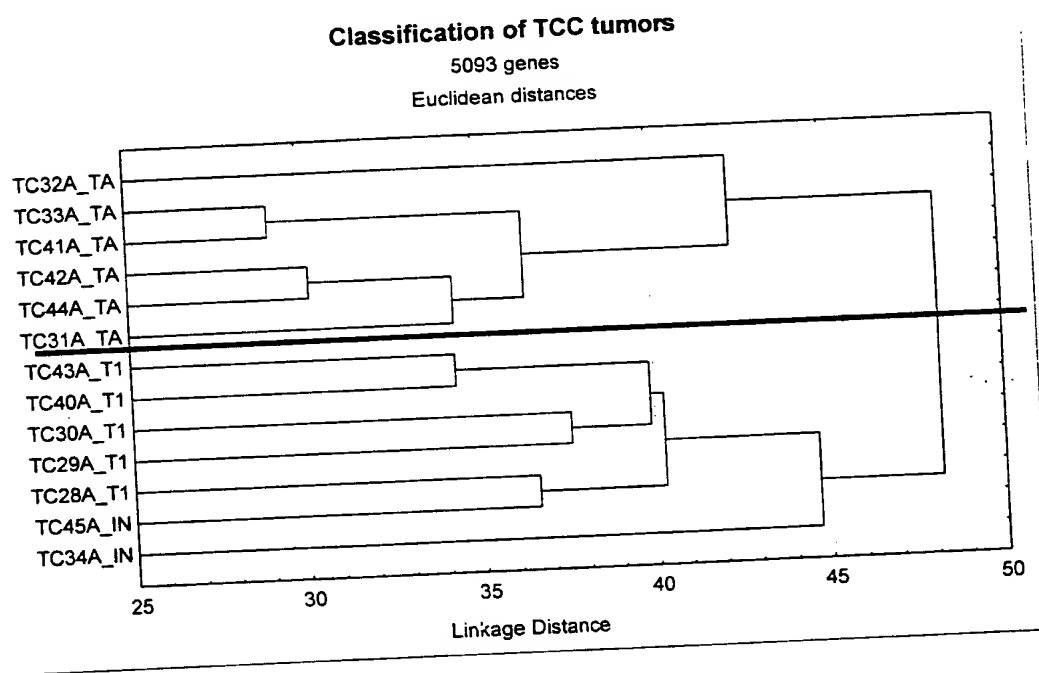
Section 6. Sequence annotation and bioinformatic analysis of the genes with unknown protein product

As shown in Table A, set 4, based on sequence annotation, 17 of the selected 22 markers are known, characterized genes; 3 additional genes code for hypothetical proteins. (One of these hypothetical proteins was identified through an EST contig of the original clone). The remaining 2 clones code for novel sequences. For one of them an EST contig was assembled, but homologous genes were not found. Only limited information is available for the 5 novel or uncharacterized genes. According to the "Secreted" probe, as described herein, one of them, the CGI-81 hypothetical protein, is a potentially secreted or membranal protein. Domain analysis also indicated the presence of a transmembranal domain in this protein. The other four genes cannot be classified with confidence, although the unknown gene in clone 70E8, can be marginally classified as "potentially secreted". No additional significant domains were identified.

Section 7. Bioinformatic analysis of known genes

According to the biomedical literature, the 17 known genes were classified into three functional groups: tumorigenesis, keratinocyte differentiation, and cell motility and proliferation. Thus, these markers fulfil a functional role in TCC. It was noted that in a previous analysis, performed with sample pools on a general-purpose human microarray, similar functional groups (and in some cases

67

the same genes) were identified, further validating the results of this study (see co-assigned patent application USSN 09/534,661 corresponding to PCT patent publication number WO 00/56935)

Only a few of these 17 markers were previously considered related to or implicated in bladder cancer. These are the keratin family of proteins, some of which are known TCC markers, midkine, for which a single report (PMID: 8653688) implies its connection to invasiveness in TCC; and FABP-5 which is related to squamous cell carcinoma of the bladder. A number of the other markers have been found to be related to other cancers, to varying extents. The specificity of markers for TCC over other cancers is an important consideration.

Four of the known up-regulated genes which are related to tumorigenesis are either membranal or secreted. Three of them were not previously reported in TCC, and the fourth (midkine) has been related to TCC invasiveness. Clearly, secreted and membranal proteins have a unique advantage for the development of a diagnostic test.

Intriguing amongst the functional group are markers related to keratinocyte differentiation. The keratins, which are cytoskeletal proteins, are known markers for keratinocyte differentiation. Five different keratins were detected in the top-scoring genes. Two of them are known to be TCC markers. The remaining three (KRT 7, 8, 17) were included in the 22 marker set. Expression of some of these and other keratins has been tested in the past by several research groups using different experimental approaches. Non-

68

consistent findings as to their regulation in TCC were reported. Thus, specific typing of the different keratins in urine of TCC patients, using a single multi-gene assay, can facilitate and improve robust TCC diagnostics.

A second major group of markers associated with keratinocyte differentiation are the S100 proteins. These are low-molecular-weight calcium-binding proteins which are probably involved in the regulation of a number of cellular processes including cell cycle progression and cell differentiation. Four different S100 proteins (S100A11, A6, A13,P) are included in the marker set. None has been previously associated with TCC. S100 proteins were implicated in other cancers (AML, colorectal). The S100P was found to be down-regulated upon androgen depletion in the androgen- dependent prostate cancer cell line LnCap. Another S100 protein, psoriasin (S100A7, not included in the set), is involved in both invasive breast cancer, squamous cell carcinoma of the bladder and psoriasis, another disease involving keratinocytes. Note, that another psoriasis-related gene, PA-FABP is also included in this proposed marker set. PA-FABP has also been implicated in squamous cell carcinoma of the bladder. Interactions between another FAPB (E-FABP) and psoriasin are well-documented in psoriatic keratinocytes. The identified S100 proteins and PA-FABP are important novel markers for TCC.

Section 8. Sequence annotation of genes with identical expression pattern in all tested samples

The diagnostic assay is based on the genes which are up-regulated in TCC. In all suggested tests, and mainly in the RT-PCR based assay, internal controls for each tested sample are beneficial. Such controls can be genes which are normally not upregulated in TCC. To this end the hybridization data was analyzed to identify genes which are:

1.   Expressed at a high, easily detectable level in all the control and TCC samples. This is important to enable detection even in small amounts of RNA.

2.   Genes which are not differential in TCC compared to normal urothelium. Such genes were specifically selected according to their normalized differentials 14 genes were detected as suitable according to these criteria (Section 9 hereunder). These genes are included in the diagnostic assays, and used as internal references for a normally, uniformly expressed gene in TCC and in non-TCC cases.

Section 9: Expression patterns, signals and annotation report for control genes

The differential expression patterns and basic annotation for the 14 non-differential genes suggested as internal controls is provided in Table E (at

70

end of specification , just before Table 1). Table E displays differential expression results, and shows signals of all genes in all hybridizations.

The expression levels are shown in the following order:

Normal urothelium (16 first hybridizations)

T1 samples (13 hybridizations)

Ta samples (10 hybridizations).

All 14 clones were fully sequenced from both sides. Full (contigized) sequences passed via the standard sequence annotation platform including sequence QC, chimer detection, and homology searches within Genbank's non-redundant genomic and non-genomic nucleotide databases, the non-redundant protein database and the EST database. EST contigs were assembled for several novel genes for which ESTs were available, and further annotated.

· The results supported the choice of any one, or a combination of two or more, of these 14 genes as internal controls.

Section 10. Class definition and characterization

Staging and grading of TCC is not straightforward. In fact, subjective decisions must often be made in order to classify tumors, and pathological experts can differ on the correct diagnosis. Unsupervised analysis methods as well as supervised methods of class prediction were used, in order to reduce the dependence on expert opinion.

71

During the quality control process (section 2), some separation between T1 and Ta samples was noted in the third set of experiments, as well as in global clustering of all 41 profiles (39 normal or non-invasive, plus 2 invasive TCC samples). Therefore clustering of the tumor samples *within the third set only* (Table D1) was pursued. This set contains the largest and most variable collection of TCC samples. A clear differentiation exists between Ta and T1 tumors. One tumor classified as T1/Ta resides within the Ta cluster, but is separated from other Ta tumors. The two invasive tumor samples are in the T1 cluster (one clearly inside, the other outlying). Thus at the level of hierarchical clustering Ta tumors are separated from T1 high-grade tumors.

Standard scoring method (example 5, section 4) was employed for class prediction in order to identify specific molecular markers that underlie the Ta/T1 separation.

The Keratin 13 gene was found to be the highest scoring gene. It is down-regulated in T1 samples and up-regulated in Ta (relative to normal samples). Keratin 13 is known to be expressed in urothelium. Its expression in urothelial tumors depends on their degree of differentiation. It is expressed only in well-differentiated tumors and absent from poorly differentiated ones (PMID: 1706547). Since most of the Ta tumors in this set were classified as low grade while the T1 tumors were mostly high grade, over-expression of the KRT13 gene can be attributed to the degree of the differentiation as well as to the staging of the tumors.

72

The expression pattern of KRT13 in all TCC hybridizations was then studied. The results are not straightforward. Approximately half of the tumors show up-regulation of KRT13, while down-regulation is observed in others, with no clear correlation to either stage or grade of the tumor. KRT13 is highly relevant to the sub-classification of TCC tumors. However, the exact correlation to the classical clinical classifications remains to be elucidated in a larger set of tumors involving different stages and grades and in follow-up studies.

EXAMPLE 5: BIOINFORMATICS ANALYSIS OF 22 SHORTLISTED GENES

1. **Sequence** – All clones were fully sequenced from both sides. Sequence passed the standard sequence criteria, except where otherwise noted (see Table 6)

2. **Annotation** – Full (contigized) sequences were passed through the standard sequence annotation platform including sequence QC, chimer detection, and homology searches to Genbank's non-redundant genomic and non-genomic nucleotide databases, the non-redundant protein database and the EST database. EST contigs were assembled for several novel genes for which ESTs were available, and further annotated. Complete annotation and sequence information is available in Table 5.

3. **Literature** – an extensive search of the literature was performed for each of the known genes. Detailed information is given below.

73

Known genes are classified according to their function in tumorigenesis, keratinocyte differentiation, and cell motility and proliferation. References are given as PMIDs.

3.1 Genes associated with tumorigensis

Genes coding for secreted and membranal proteins

3.1.1. Syndecan 1 (Accession: gi|4506858|ref|NM_002997.1|)

An integral membrane protein, 310 amino acid-long, with a signal peptide at its NH$_2$-terminus. Contains a matrix-interacting ectodomain with putative glycosaminoglycan attachment sites, a hydrophobic membrane-spanning domain and a cytoplasmic domain. Is connected to cell aggregation in malignant mesotheliomas with epithelial and/or sarcomatous morphology and is required for wnt-1-induced mammary tumorigenesis in mice. On the other hand, its expression is inversely correlated to the aggressiveness of basal cell carcinoma. **PMID: 2324102, 10912783, 10888884, 10770430.**

3.1.2. Hepatocyte growth factor activator inhibitor type 2 (HAI-2) (Accession: gi|2924619|dbj|AB006534.1|AB006534)

HAI-2 is a Kunitz-type serine protease inhibitor which was recently identified as a potent inhibitor of hepatocyte growth factor activator. It was also independently reported as placental bikunin (PB) and as a protein over-expressed in pancreatic cancer. However, its expression was conserved in the neoplastic colorectal mucosa, and no relationship was found between HAI-2/PB mRNA levels and colorectal tumor stages. HAI-2 is produced in a membrane-

74

associated form and secreted in a proteolytically truncated form. PMID: 10695988 , 10762618

3.1.3. Midkine (neurite growth promoting factor) (Accession: gi|4505134|ref|NM_002391.1|)

Midkine is a heparin-binding growth factor, implicated in various biological phenomena such as neuronal survival and differentiation, tissue remodeling and carcinogenesis. In the G401 cell line, midkine initiates a cascade of intracellular protein tyrosine phosphorylation mediated by the JAK/STAT pathway after binding to its high affinity p200(+)/MKR cell surface receptor. The most intriguing feature of midkine in cancer is its augmented expression in advanced tumors at a very high frequency in a non-tissue specific manner. In addition, its high expression is also detected in precancerous lesions. Midkine exerts carcinogenesis-related activities, including transforming, anti-apoptotic, angiogenic and fibrinolytic ones. These data provide a possibility of clinical application of midkine. Serum midkine level can be a useful tumor marker. Gene therapy using its promoter region and therapeutic strategy choosing midkine as a molecular target were also suggested. MK was suggested as a marker for early and latent bladder cancer disease (specificity of 0.86). Recent publication demonstrated good correlation of MK over-expression with poor outcome in patients with invasive cancers. PMID: 10879061, 8714367, 10902971, 10626184, 10545795, 10408712

3.1.4. Solute carrier family 2 SLC2A1 (GLUT1) (Accession: gi|5730050|ref|NM_006516.1|)

Increased expression of glucose transporter1 (GLUT1) has been reported in many human cancers. Suppression of GLUT1 mRNA has been shown to suppress tumor growth. Some studies have reported associations between its expression and proliferative indices, whilst others suggest that GLUT1 can be of prognostic significance, especially in lung cancer. No connection between GLUT1 up-regulation and TCC has yet been reported. PMID: 10983690, 10806305, 10795374.

Genes coding for intracellular proteins

3.1.5. Cystatin B (Accession: gi|7263011|gb|AF208234.1|AF208234)

Cystatins are endogenous inhibitors of lysosomal cysteine proteinases, the cathepsins (Cats). Imbalance between cathepsins and cystatins, associated with metastatic tumor cell phenotype, can facilitate tumor cell invasion and metastasis. Cystatins were found to be up-regulated in relation to inflamation and cancer (breast, lung, brain and head and neck tumors, and in body fluids of ovarian, uterine, melanoma and colorectal carcinoma). In contrast, reduced expression of cystatin B was found in esophageal-carcinoma tissue and was associated with lymph-node metastasis. The application of cystatins for prognosis, diagnosis, follow-up and anticancer therapy has been proposed (but not for TCC). In the preliminary experiments in TCC, using general microarray

76

containing 10,000 human ESTs, Cystatin A was found to be up-regulated in TCC pool compared to pool of normals. **PMID: 10566975, 9769367, 9583733,10514828.**

3.1.5 Opa-interacting protein OIP3

(Accession:gi|2815605|gb|AF025439.1|AF025439)

Opa proteins are a family of outer membrane proteins involved in gonococcal adherence to and invasion of human cells. Pyruvate kinase M2 is OIP3 which binds to OPA proteins. Modulation of type M2 pyruvate kinase activity by the human papillomavirus type 16 E7 oncoprotein has been demonstrated. PMID: 9990017, 9692838

3.2 Genes associated with abnormal differentiation of keratinocytes

3.2.1. Keratins: keratin 19, keratin 7, keratin 8, keratin 18, keratin 17

Keratins, or cytokeratins, represent a family of more than 20 different polypeptides which are important markers of epithelial cell differentiation. Both gene expression and protein levels are elevated (and even used as a marker) in several pathological conditions including breast cancer, kidney tumors, small cell lung cancer (SCLC), and pre-eclampsia. Measurements of cytokeratins 19 and 20 levels in serum and urine are in use as tumor marker for bladder cancers.

Keratin 18 is a type-I keratin that is found in a variety of simple epithelial tissues. Pancreatic exocrine acinar cells and endocrine islet cells are well-differentiated cells which express the keratin combination 8 and 18, whereas the less-differentiated cells of the ductal tree are characterized by the additional expression of keratin 7, keratin 19, and, in the rat, keratin 20. Levels of keratin 7 and 20 are increased in rectal adenocarcinoma and Paget's disease. PMID: 10755601, 10707834, 10782894, 10775728, 10762743, 9614373, 8911513, 9445193, 2434380

3.2.2. S100 proteins: S100A11, S100P, S100 Calcium binding protein A13, S100A6

S100 proteins are low-molecular-weight calcium-binding proteins of the EF-hand superfamily and appear to be involved in the regulation of a number of cellular processes such as cell cycle progression and differentiation. More than 10 members of the S100 protein family have been described from human sources so far. Induced expression in tumors of some of these genes has been reported.

S100A11 (or S100C/ Calgizzarin)

Calgizzarin is a nuclear protein which inhibits the actin-activated myosin Mg(2+)-ATPase activity of smooth muscle in a dose-dependent manner. Other Ca(2+)-binding proteins such as S100A1, S100A2, S100B, and calmodulin do not inhibit actin-activated myosin Mg(2+)-ATPase activity. Calgizzarin can be involved in the regulation of actin-activated myosin Mg(2+)-ATPase activity

78

through its Ca(2+)-dependent interaction with actin filaments. It is expressed in most tissues and cell lines, and co-localized with the psoriasin gene S100A7 and other S100 genes to human chromosome 1q21-q22.

Calgizzarin was found to be remarkably elevated in colorectal cancers compared with that in normal colorectal mucosa. No similar alteration in expression was detected in breast cancer. PMID: 10486266, 10623577, 7591220, 7889529

S100P (Accession: gi|5174662|ref|NM_005980.1|)

S100P overexpression is an early event that might play an important role in the immortalization of human breast epithelial cells in vitro and tumor progression in vivo. S100P expression was downregulated after removal of androgen from LnCAP prostate cancer cell line. PMID: 10639564, 8977631

S100 Calcium binding protein A13

(Accession: gi|5174658|ref|NM_005979.1|)

S100A13 was found to be widely expressed in various types of tissues including skeletal muscle, heart, kidney, ovary, small intestine and pancreas. It was shown to bind anti-allergic drugs and thus to be involved in the inhibition of degranulation of mast cells. Also, it was shown to be involved in the regulation of FGF1 activity. PMID: 10722710, 8878558, 9712836, 10051426

Growth factor inducible 2a9/calcyclin/S100A6 (Accession: M14300)

2A9 was isolated from stimulated quiescent fibroblasts. It is induced by growth factors and over-expressed in AML. S100A6 was also suggested to be involved in the progression and invasive process of human colorectal adenocarcinomas. PMID: 10656447, 1952954

3.2.3 PA-FABP - Fatty acid binding protein 5 (psoriasis-associated)

(Accession: gi|4557580|ref|NM_001444.1|)

The fatty acid-binding protein (FABP) family consists of small, cytosolic proteins believed to be involved in the uptake, transport, and solubilization of their hydrophobic ligands. PA-FBP can be involved in keratinocyte differentiation. In normal skin, PA-FABP is expressed in basal and prickle cell layers, and more strongly in the granular cell layer. In psoriatic skin, PA-FABP is expressed in suprabasal layers and more strongly in more differentiated keratinocytes. In squamous cell carcinoma, PA-FABP shows very strong expression in squamous nests. Serum levels of intestinal fatty acid-binding protein (I-FABP) serve as diagnostic marker for mesenteric infarction (acute ischemic diseases of the bowel). Expression of PA-FABP has been linked to squamous cell carcinoma of the bladder. PMID: 9521644, 9438903, 8566578, 8092987, 9307301

3.3 Genes involved in cell motility and proliferation

3.3.1 Actin gamma 1 (Accession: gi|4501886|ref|NM_001614.1| gi|4501886|ref|NM_001614.1|)

Ubiquitously expressed in all eukaryotic cells. Beta and gamma actins co-exist in most cell types as components of the cytoskeleton and as mediators of internal cell motility.

3.3.2 37 kD laminin receptor precursor/p40 ribosome associated protein

(Accession: HSU43901)

The 37 kD precursor of the 67 kD laminin receptor (37LRP) is a polypeptide whose expression is consistently up-regulated in aggressive carcinoma. Interestingly, the 37LRP appears to be a multifunctional protein involved in the translational machinery and has also been identified as p40 ribosome-associated protein. It is distributed on the cell surface as laminin binding protein p67 (LBP/p67), in the nucleus, and on 40S ribosomes. PMID: 8760291, 10079194

Throughout this application, various publications, including United States patents, are referenced by author and year and patents by number. Full citations for the publications are listed below. The disclosures of these publications and patents in their entireties are hereby incorporated by reference

into this application in order to more fully describe the state of the art to which this invention pertains.

The invention has been described in an illustrative manner, and it is to be understood that the terminology which has been used is intended to be in the nature of words of description rather than of limitation.

Obviously, many modifications and variations of the present invention are possible in light of the above teachings. It is, therefore, to be understood that within the scope of the described invention, the invention can be practiced otherwise than as specifically described.

82

Table E

Index	GeneDescription1	Acc1	GeneID	TC7	TC8	TC9	TC1	TC1	TC2	TC2	TC2	TC2	TC4	TC4	TC3	TC3	TC3	TC4	TC4	TC3	TC3
1	Human 54 kDa protein mRNA, complete cds ; nt_non_genomic(ident	gi|407307|gb|U02493.1||	TCC-29E_6	-1	-1.2				-1.2	-1						-1		-1.1		-1.5
2	none:11_TCC_74D3_1_Mi3F_fa none:25_TCC_74D3_1_Mi3R.fa		TCC-74D_3							-1.3									-2.1	-1.8
3	(AL121588) dJ477O4.1 (novel protein similar to otoferlin and dysferlin	gi|7871862|emb|CAB89	TCC-70E_2	-1.2	-1.1	-1.2	-1		-1	-1.1	-1.4	-1.2	-1.2	-1.1	-1.3	-1.3	-1.2		-2.1	-1.9
4	Homo sapiens full length insert cDNA clone ZC48G12 ; nt_non_genom	gi|3483555|gb|AF08621	TCC-76B_7	-1.1	-1.1	-1.1			-1.2	-1.1	-1.3	-1.1	-1.1		-1.3	-1.3	-1.1		-1.1	-1.7
5	Homo sapiens GPI-anchored metastasis-associated protein homolog	gi|7656943|ref|NM_0144	TCC-89H_5	-1	-1.1	-1	-1	-1	-1.2	-1.2	-1.2	-1.1	-1	-1	-1.2	-1.3	-1.2	-1.2	-1.2	-1.2
6	Homo sapiens cDNA: FLJ22720 fis, clone HSI14320 ; nt_non_genom	gi|10439219|dbj|AK0263	TCC-101D_4	-1	-1.2	-1.2		-1	-1.1	-1.1	-1.2			-1	-1.2	-1.3	-1	-1.1	-1.2	-1.8
7	low molecular mass ubiquinone-binding protein	Homo sapiens]|UCR(	gi|7657488|ref|NP_0552	TCC-67C_11	-1.1	-1	-1.2	-1.1		-1.2		-1.2		-1.1	-1.2	-1.2	-1.3	-1.1	-1.3		-1.1
8	none:14_TCC_79B1_1_Mi3F_fa Homo sapiens quiescin Q6 (QSCN6	gi|4506360|ref|NM_002	TCC-79B_1	-1.1	-1.2	-1.3	-1.3	-1.2	-1.4	-1.1	-1.4			-1.3	-1.3	-1.3	-1.2	-1.3		
9	Homo sapiens mRNA; cDNA DKFZp564C2282 (from clone DKFZp56	gi	7328021	emb	AL1619	TCC-56G_1	-1.1	-1.3	-1.3	-1.2	-1.4	-1.1	-1.1			-1.2	-1.2	-1.1	-1.2			-1.1
10	Homo sapiens full length insert cDNA clone ZC48G12 ; nt_non_genom	gi|3483555|gb|AF08621	TCC-47D_9	-1.1	-1	-1	-1.1	-1	-1	-1	-1	-1.1	-1.1	-1.2	-1.1	-1.2	-1.1	-1.3	-1.2	-1.2
11			TCC-60B_2	-1.1	-1.4	-1.2	-1.2			-1.2					-1.2	-1.2	-1.3	-1.3	-1.1
12	NADH dehydrogenase (ubiquinone) 1 subcomplex unknown, 2 (14.5l	gi|4758784|ref|NP_0045	TCC-23D_4			-1.1			-1			-1			-1	-1.1	-1	-1.3		
			TCC-20F_5	-1.2	-1.1	-1.1	-1.2	-1.1	-1.1	-1.2	-1.2		-1.2				-1.1	-1	-1.2	-1.6
13	——																			
14	——																			

83

Table E

TABLE 2

88

TABLE III

>40_TCC_13F11_M13F.fa   TIME: Sun Sep 10 11:42:06 2000 trimming
information: raw_sequence:582 (high quality:29-320) sequence:97-
252 [length:156]
TCCGTCTCATTGAGGGTCCTGAGGAAGTTGATCTCATCATTCAGGGCATC
CACCTTGGCCTCCAGCTCCACCTTGCTCATGTAGGCAGCATCCACATCCT
TCTTCAGCACCACAAACTCATTCTCAGCAGCTGTGCGGCGGTTAATTTCA
TCTTCG

>04_TCC_94G3_M13F.TXT.fa ,constant: 15, poly A: yes
AAGGCTTATTCCATCCGGACCGCATCCGCCAGTCGCAGGAGTGCCCGCGACTGAGCCGCC
TCCCACCACTCCACTCCTCCAGCCACCACCCACAATCACAAGAAGATTCCCACCCCTGCC
TCCCATGCCTGGTCCCAAGACAGTGAGACAGTCTGGAAAGTGATGTCAGAATAGCTTCCA
ATAAAGCAGCCTCATTCTGAGGCCTGAGTGAAAAAAAAA

>20_TCC_60H4_M13F.TXT.fa ,constant: -1, poly A: no
CANTATATAACNAATTGGAGCTCAATNGCNCGCGGNCGCGTGTCTTCTGGGTAGAGGGAT
GNGAAGGAAGGGACCGTTACCCCCGGCTCTTCTCCTGACCTGCCAATAAAAATTTATGGT
CCAAGGNAAAANA

>26_TCC_44C1_M13F.TXT.fa ,constant: -1, poly A: no
ACTCATTGAACTTGAGCTCCGANTCCTGATTCNCATCNAAGCTCTNNATCTGCTCATCAN
GAGANCCCACATCCTTGAGCAGATGGNGCANCTGCTGNTTAACCANCTCTNNGAACTCGN
AGANNNTAAGGCTATCCTTCCGGNCCTCCTGCCTTGCAAAGGTGAAGAAAGTGGTGNNCA
CNGTCNCAATGGANTCCTCTAGCTCTGTCAGTGGTTCTGCTGCNATTATGGAACCTGAGG
CCAAAGCTGATGTCCTCAAGGGGCTAGCTGACCTTTGTCAGGGCTGACCTCTCCTCAGCG
GCAGCAGGGCAGAGTGCTGAACCCAGGAACCCACAGATCCTCCCCGNTCCTGTCTCCCGG
TGACAAGGGTCCTGGAACGGCGCGTCTCTGACTCCCTGCTCCAGGACGGGTTTAAGT

>29_TCC_48G1_M13F.TXT.fa ,constant: -1, poly A: no
ACTTTGAGAAGGCAGGACTCAAATGATGCCCTGGAGATGTCACAGATTCCTGGCAGAGCC
ATGGTCCCAGGCTTCCCAAAAGTGTTTGTTGGCAATTATTCCCCTAGGCTGAGCCTGCTC
ATGT

>31_TCC_65B9_M13F.TXT.fa ,constant: -1, poly A: yes
GACTAGAACCCACCCCCTTNCCTTCCAGCCTTTCTGTCATCATCTCCACAGNCCANCCAT
CCCCTGAGCACACTAACCATCTCATGCAGGCCCCACCTGCCAATAGTAATAAAGCAATGT
CACTTTGTTAAAACATGAAAAAAAAA

>47_TCC_91B11_M13F.TXT.fa ,constant: -1, poly A: yes
CTAGTATACACTCCNCATAGNATACGTTGCAGCTCAATTGCGCGCGGNCGCGGACGACGA
CCTGCGAGGGTGTCTTCTGGGTAGAGGGATGGGAAGGAAGGGACCCTTACCCCCGGCTCT
TCTCCTGACCTGCCAATAAAAATTTATGGTCCAAGGAAAAAAAAA

>10_TCC_53H11_T3
TTTTTTNNATNTTATTTTGGGTATTGGTGTTNTTTCTTTTTTTCCTCTTNCCTTCTTAACT
CAAGACTTGTAGTGTTGTAAACCTGCCTCACAAAATACATGGTAATAACTTNTCTTTAAA
AAAANAAAAAAGACAGNCTTNACACCATTTCTAATNGNANNACTATTTTTGGGCAATGTT
ATGCACCACTTCAATTTCCCCATTGTGACCCCTATCACTTCATTTGATATCCCTTTTTNGA
CCCANCCATCTCCTTCATATATGGGCATGTCCATAGATTGACAAAGAAAGTTTACACTTT
NGAATAAAGATGCAAAGTATGCAAAAACATTAATACTGATGCNAAAAAAAANTANAAAAA

>07_TCC_57B3_M13F.TXT ,constant: -1, poly A: yes
GGTACCGACGGACCTGCGGAGACTCCTGCCCTGTTGTGTATAGATGCAAGATATTTATAT
ATATTTTTGGTTGCAATATTAAATACAGACACTAAGTTATAGTATATCTGGCAAGCCAAC
TTGTAAATCACCACCTCACTCCTGTACTTACCTAAACAGATATAAATGGCTGGTTTTTAA
GAAAAAAAAA

>11_TCC_25F2_M13F.TXT ,constant: -1, poly A: no
ACCCTGGGAGAGAAGTTTGAAGAAACCACAGCTGATGGCAGAAAAACTCAGACTGCTGCA
ACTTTACAGATGGTGCATTGNGTCAGCATAGGAGTGAGATGGGGAAGGAAAGCACANTAA
CAAGAAAATTGANAGATGNTAAATTAGTGNTGGAGTGTGTCATGAACAATGCACCTGT

>25_TCC_50G5_M13F.TXT ,constant: 17, poly A: yes
TAGTGTGGAAGCATAGTGAACACACTGATTAGGTTATGGTTTAATGTTACAACAACTATT
TTTTAAGAAAAACATGTTTTAGAAATTTGGTTTCAAGTGACATGTGTGAAAACAATATCG
ATACTACCATAGTGAGCCATGATTTTCTAAAAAAAAA

>26_TCC_50G6_M13F.TXT ,constant: 17, poly A: yes
TAGTGTGGAAGCATAGTGAACACACTGATTAGGTTATGGTTTAATGTTACAACAACTATT
TTTTAAGAAAAACAAGTTTTAGAAATTTGGTTTCAAGTGACATGTGTGAAAACAATATTGT
ATACTACCATAGTGAGCCATGATTTTCTAAAAAAAAA

>26_TCC_75E3_M13F_B04_032.ab1.TXT ,constant: 16, poly A: yes
AAAGAGGGCGGCAGGGGCCTGGAGATCCTCCTGCAGACCACGCCCGTCCTGCCTGTGGCG
CCGTCTCCAGGGGCTGCTTCCTCCTGGAAATTGACGAGGGGTGTCTTGGGCAGAGCTGGC
TCTGAGCCGCCCTCCATCCAAGGCCAGGTTCTCCGTTAGCTCCTGTGGCCCCACCCTGGG
CCCTGGGCTGGAATCAGGAATATTTTCCAAAGAGTGATAGTCTTTTTGCTTTTTGGCAAA
ACTCTACTTAATCCAATGGGTTTTTCTCTGTACAGTAGATTTTCCAAATGTAATAAACTT
TAATATAAAGTAAAAAAAAA

>30_TCC_76B3_M13F_F04_042.ab1.TXT ,constant: 16, poly A: yes
AAAGTCATCCTCCGTCTACCAGAGCGTGCACTTGTGATCCTAAAATAAGCTTCATCTCCG
GGCTGTGCCCCTTGGGGTGCAAGGGGCAGGATTCTGCAGCTGCTTTTGCATTTCTCTTCC
TAAATTTCATTGTGTTGATTTCTTTCCTTCCCAATAGGTGATCTTAATTACTTTCAGAAT
ATTTTCAAAATAGATATATTTTTAAAATCCTTAAAAAAAAA

>38_TCC_56E11_M13F.TXT ,constant: -1, poly A: yes
CTCTCCAGTTTGCACCTGTCCCCACCCTCCACTCAGCTGTCCTGCAGCAAACACTCCACC
CTCCACCTTCCATTTTCCCCCACTACTGCAGCACCTCCAGGCCTGTTGCTATAGAGCCTA
CCTGATGTCAATAAACAACAGCTGAAGCAAAAAAAAA

>46_TCC_78B11_M13F_F06_058.ab1.TXT ,constant: 16, poly A: yes
AGGAAAGGTGNGNGCTGGAAGCACTGAACCTACCTCATCCTCCTGGTGGGTGTGGCTACC
CTCGCCACCCCAAATTCCATGTCATTAAAGAACAGCTAAATTCAAAAAAAAA

>53_TCC_79G2_M13F_E07_054.ab1.TXT ,constant: 16, poly A: no
TGTCCGTCTTCACCCATCCCCAAGCCTACTAGAGCAAGAAACCAGTTGTAATATAAAATG
CACTGCCCTACTGTTGGTATGACTACCGTTACCTACTGTTGTCATTGTTATTACAGCTAT
GGCCACTATTATTAAAGAGCTGTGTAACATCAAAAAAA

>82_TCC_89G3_M13F_B11_092.ab1.TXT ,constant: 16, poly A: yes
CAGGAGACCATCCGCGTCACCAAGCCCTGCACCCCCAAGACCAAAGCAAAGGCCAAAGCC
AAGAAAGGGAAGGGAAAGGACTAGACGCCAAGCCTGGATGCCAAGGAGCCCCTGGTGTCA
CATGGGGCCTGGCCCACGCCCTCCCTCTCCCAGGCCCGAGATGTGACCCACCAGTGCCTT
CTGTCTGCTCGTTAGCTTTAATCAATCATGCCCTGCCTTGTCCCTCTCACTCCCCAGCCC
CACCCCTAAGTGCCCAAAGTGGGCAGGGACAAGGGATTCTGGGAAGCTTGAGCCTCCCCC
AAAGCAATGTGAGTCCCAGAGCCGGCTTTTGTTCTTCCCCACAATTCCATTACTAAGAAA
CACATCAAATAAACTGACTTTTTTCCCCCAAAAAAAAA

>35_TCC_21D6_M13F_C06_037.ab1.fa   TIME: Wed Aug 9 12:48:31 2000
trimming information: raw_sequence:889 (high quality:34-340)
sequence:95-456 [length:362]
CTTTGACGTGGAGAGGAACTCCTGCAATAACTTCATCTATGGAGGCTGCC
GGGGCAATAAGAACAGCTACCGCTCTGAGGAGGCCTGCATGCTCCGCTGC
TTCCGCCAGCAGGAGAATCCTCCCCTGCCCCTTGGCTCAAAGGTGGTGCT
TCTGGCGGGGCTGTTCGTGATGGTGTTGATCCTCTTCCTGGGAGCCTCCA
TGGTCTACCTGATCCGGGTGGCACGGAGGAACCAGGAGCGTGCCCTGCGC
ACCGTCTGGAGCTCCGNAGATGACAAGGAGCAGCTGGTGAAGAACACATA
TGTCCTGTGACCGCCCTGTCGCCAAGAGGACTGGNGAAAGGGAGGGGAGA
CTATGTGTGAGC

>46_TCC_27H5_M13F_F06_058.ab1.fa   TIME: Wed Aug 9 12:48:35 2000
trimming information: raw_sequence:892 (high quality:169-406)
sequence:170-287 [length:118]
AAAAAGAGTAAAACACTTTCAGTTTCTCCCCTTTAGCCCCTAAAACAACA
TCTTACAGTCTGGATCTGGATCTACCTATACAGTCCTACATTAGCTTCTA
AAATATTTGTCAGGAGGG

TABLE IV

>31_TCC_10E8_M13F.fa   TIME: Sun Sep 10 11:42:01 2000 trimming
information: raw_sequence:549 (high quality:25-313) sequence:98-
313 [length:216]
CCCAAATGGAATGTTGCCCCCCTTAAACACCATTTTCCCTCCAGGACCACC
TTGGTTTCTAGGCACTGTGGTTCTTGGCAGGGGCTGTCTTAGGTAAAAGG
GTAGTTGTGGAGCTACAGTCTGAAGAACATAGCTTGGGCTCAAGTTCAAA
TGAGCCATCTTTTTCCTTTGCGTTTTTCTTGACTGAAGGTGAGATGTTAT
TTGTGGCATGTGAACT

>09_TCC_101C11_M13F.TXT.fa ,constant: 16, poly A: yes
ACAAAGACTGCTGATAACTATCTGTGATTGATAGGAAATTTTTTTTCTTGATTTCTCTGT
GAGAAATGTAATGCTGACTTTTATAAAGCCTGGACTTCTACTTTATTTAATAAATCAATG
TTTGCAATGGTAAAAAAAAA

>11_TCC_101E11_M13F.TXT.fa ,constant: 15, poly A: yes
GCAATAAAGCTGTCCATTCAATTCCAAATACTGGTTTTAAGNGTATAGCCACTGATATTC
TTTCATGTNTAGAAATTCTTTCTGTTATTATTCAAGAAAATGTTTTTAATCATGCTAATA
AACTTTTTTGCAGATGAAAAAAAAA

>15_TCC_57C3_M13F.TXT.fa ,constant: -1, poly A: no
GGNACCACGTACCTGCTGAATGTNTCNNCGNNATGNCGNCAGGCCATGCTGTTGCTGATN
TANTACTNTGAAAATANGGATATCATGATGGGAATGCATGTCATGAGGTCCAGANTCGTT
CTACTGTCNATAANCTGTNTACTNGCGTTGANAANAAANGATGTCAAAGNCCCCCCGTAA
AAANGTA

>44_TCC_70E8_M13F.TXT.fa ,constant: 15, poly A: yes
GTCCCAGTCTTCACCAGGTGTCTCTCCTCTTTTACTCAGGAGGACTTTCCCAGGAAAACC
ATGCCACTAGCAAAAAAAAA

>03_TCC_57E11_T7.TXT.fa ,constant: 16, poly A: yes
TGAGTGTCTTCAGGCCAACCTGGTGGAAATGTTGTTCTCTGAAGATTAAGATTTTAGGAT
GGCAATCATGTCTTGATGTCCTGATTTGTTCTAGTATCAATAAACTGTATACTTGCTTTG
AATTCATGTTAGCAATAAATGATGTTAAAAAAAAA

>08_TCC_70E7_M13F_H01_015.ab1.TXT ,constant: -1, poly A: no
GGATCGACGACCTGCTTCCCAGANGCGNNCNNGAGGNCCNCTTGTTNNNGNCNNGNANAC
NNACCCANTTNANTTNNAGCCTTTNTGNAATAAATATACACAGGCCACCCATGCCNTGAG
CACACTAACCACNTGATGCAGGCCCCACCTTGCCAATAGTAATAAAGCANTGGGACGTTT
TTTA

>13_TCC_71E4_M13F_E02_018.ab1.TXT ,constant: -1, poly A: no
GGGCCAAAGCCCGNGCATCCAANCCCANGCAAGGNACAAANGANCNNGGAGAGGANNACC
CAAGCANNTNNCAACCATCAAATGGAGGGCANGCCCGGGG

>15_TCC_71H8_M13F_G02_019.ab1.TXT ,constant: -1, poly A: no
GGGCCAAAGCCGNGCATCCAANCCCANCGCANGGNANAAANGANGANGGANANGGATNAC
CCANGCCTNTATTAACCATCAANTGGGANGGCAAGCCCGGGGCATNTATTGATT

>21_TCC_43E2_M13F.TXT ,constant: -1, poly A: no
AGGACCCCTGAANACNACACAGATCTGTGNGAAACAANGGNACNTAGCGTCCCNAAAGTG
CCNGGTTNNNGTANNCNNAGNGNGNGACCNGNGCNCATNT

>24_TCC_36C7_M13F.TXT ,constant: 16, poly A: yes
ATCCAGAGACCATCAATCCTGCTAGAGTGCAGGGTGGCAAGCACCCAAGGGTGGCTGACC
AAGACTGCAGAGTCTCCTCCATCTTCAGGTCCATTCAGCCTCCTGGCATTTAACTACCAG
CATCCAGTGGTCCCCAAGGAATCCCTTCCTAGCCTCCTGACATGAGTCTGCTGGAAAGAG
CATCCAAACAAACAAGTAATAAATAAATAAATAAACTCAAAAAAAA

>57_TCC_80C9_M13F_A08_056.abl.TXT ,constant: 17, poly A: yes
CTGCAGGAGTCAGCGTTCAATCTTGACCTTGAAGATGGGAAGGATGTTCTTTTTACGTAC
CAATTCTTTTGTCTTTTGATATTAAAAAGAAGTACATGTTCATTGTAGAGAATTTGGAAA
CTGTAGAAGAGAATCAAGAAGAAAAATAAAAATCAGCTGTTGTAATCACCTAGCAAAAAA
AAA

>14_TCC_9B6_M13F_F02_026.abl.fa   TIME: Wed Aug 9 12:48:25 2000
trimming information: raw_sequence:871 (high quality:73-413)
sequence:98-394 [length:297]
CACGCATATGGGGCCAGTTCCACATATTTGGCAACCAGACCAGCATCCAG
GACAACACAAAGTATGTTGTTTGTTGTTAGAGGGCTTGGGACATTTCACT
CTTTGCCAGCCTCAGCTTAATCCAGGAGACAAAGATTATTTTCCTTATTA
TCTCTTCTGCATAGGATCTGCAATCAGAACTATTGAACTTCTCCATTCAG
ACCGCCACTCACACCTATGGGAAAAGGGTAATGTATCATCGGCTTAGCAA
CAGGGAATACTATTCGTATGATGGAAAATGGGGACAAAAGGCTTTGG

>24_TCC_12F3_M13F_H03_031.abl.fa   TIME: Wed Aug 9 12:48:28 2000
trimming information: raw_sequence:842 (high quality:82-340)
sequence:98-476 [length:379]
CTATGAATAGCTTCTTGCTTTATGACTTTAGGATTAACTTGTAAAAAACA
TATCCTGAACTAAGATATGCAAAATACTCATTTTCAAGTTATGGAAATGT
GTTTGTGGCATATAGGACTGTGGGGTCTGTGTGTGTAGTGAGAGTGTGTA
TCCACTATTATAACTGGAATTTAATTTACATTCATAAACTACTATATTTC
CCATCTTGCAAATCATTTTATGTCTCATCTGTTTTTCCTTTCGGNTATAT
CTTTGGNTTTGAATACCAACATTTAAAATGATGGNATTTTATCTTTTAAA
CTTAAAAATTATTTAATACAGCTATATGGACCTTATAAAATTGATTTCTT
ATTTATTATTAGACATTACTACTAAAAGG

>26_TCC_13H10_M13F_B04_032.abl.fa   TIME: Wed Aug 9 12:48:29 2000
trimming information: raw_sequence:874 (high quality:67-356)
sequence:99-261 [length:163]
CTAACCCACGATTCTGAGCCCTGAGTATGCCTGGACATTGATGCTAACAT
GACCATGCTTGGGATGTCTCTAGCTGGTCTGGGGATAGCTGGAGCACTTA
CTCAGGTGGCTGGTGAAATGACACCTACGAAGGAATGAGTGCTATAGAGA
GGAGAGAGGAGTG

>28_TCC_16D12_M13F_D04_041.abl.fa   TIME: Wed Aug 9 12:48:29 2000
trimming information: raw_sequence:866 (high quality:71-411)
sequence:95-602 [length:508]
CAGCTGATGTCATGTGGTGCTGAGAAGAAAGCAGATCACACTTCATCACA
GAAAGAATGCCTTGTGATTATCTTCTCCACATCTGAAATTCCTTTTGACA
CCTGCATTGGGCCGACTGCCATTCCCATGACTGCTGCACCTGCGTTTTTA
GAGAATGCCTCATAACCCACTGATTCTCATTCACAGAGAATGGGAATACG
GAATGAAGAAAGATTCCAGCAGCTTATAGAAGGATAGCAATATTTTGGGA
CAGGGAAAATCCTGTCATACCTCACCTCTTCCTCAGGAGGAGTTCTGAGC
TGGTCCTGCTTTTCATAGNTGTTTCTTTTCTTCCACTTAAGAACTCATAG
ATTTTTCTTACTGTCCTAAGGAAGTCCTTACCTCTGAGGTATCTCCTCAA
TGAATACTGTTTTCAAGGCTGAAATAGTTCATTATGTTAATAACCTTCTT
TATGTTCTCAGGGAAATGCTTAGGTGGTGTCACAAAAAGGGCCTTTTCTT
TNCTTTNC

>29_TCC_17A5_M13F_E04_034.abl.fa TIME: Wed Aug 9 12:48:30 2000
trimming information: raw_sequence:861 (high quality:83-477)
sequence:99-187 [length:89]
CTTCAAAAAGTGTATTGTCAAACATACCTAACTTTCTTGCAATAAATGCA
AAAGAAACTGGAACTTGACAATTATAAATAGTAATAGTG

>54_TCC_30E5_M13F_F07_062.abl.fa TIME: Wed Aug 9 12:48:37 2000
trimming information: raw_sequence:836 (high quality:65-394)
sequence:90-235 [length:146]
CAATTTGTTATAGTATAGTATCAAATTTCTATATAGATTTTATACCTCAG
TGGGGAAAAATAACTGATTCCAATGACATTCATTTTGTTTTCATCTGTGA
TAGTCATGGATGCTTTTATTTTCCTTGGGGTGCTGAAATTGAGCTG

>59_TCC_34D5_M13F_C08_065.abl.fa TIME: Wed Aug 9 12:48:39 2000
trimming information: raw_sequence:875 (high quality:63-434)
sequence:96-244 [length:149]
CCTGCCAAAATCCTACCACAGGATAACATTACAAGCAAAAAATTTACATG
TTCCAAAGTCTACCACACTCAAGAAGTTACTAAGAACTCTTGCAGAATAA
AAGTCACCATTTTAGAAATGCAAACCCACTTCCAACCTTTGCACAGTCC

>72_TCC_37E11_M13F_H09_079.abl.fa TIME: Wed Aug 9 12:48:43 2000
trimming information: raw_sequence:899 (high quality:35-432)
sequence:97-444 [length:348]
CATTTTTAGTGACATTTTAAAAGCAGTCAGATTCTATAAATGGCAAGTAA
GCCTGAAGTGAGGATACTGCAATTTTCGGGAGAAAAGAACAGCAGCTCTTT
AAGTGTTTGCATTTTCTATTTGGGGGGCAGGGAACTGTCATTCATTTTGC
ACAATTCTTGAACTGATGTCAGCACCCGAGTGGCTCCTGAATTTAAGTCT
GGGACGACATCTTTTATTTTTACATGAATCTTTAAACAATTCTGTGAGCA
AAGTTTGTAGCTGCTGGATTATTGTCTGTCTTTATAGCAAGTTCCAGTAA
ACCACAAGTATGGCAAAGCTTATCCAATTTTATGCTTGNAGCAGTCAG

94

Table 5 : Subset of 22 genes identified as potential TCC markers

Gene ID	GENE DESCRIPTION	2nd-ary Gene Description	Accession Number	A
TCC-75E3	Sequence 202 from Patent WO9947669 ; nt_non_genomic(identity):contig_TCC_75E3_RF.fa	H.sapiens syndecan-1 gene (exons 2-5)	gi|10042244|emb|A X017423.1|AX0174 23	
TCC-71E4	Sequence 102 from Patent WO9953040 ; nt_non_genomic(identity):contig_TCC_71E4_RF.fa	Homo sapiens mRNA for hepatocyte growth factor activator inhibitor type 2, complete cds	gi|10041170|emb|A X014903.1|AX0149 03	x
TCC-94G3	Human mRNA fragment for mesothelial type II kerat ; nt_non_genomic(identity):contig_TCC_94G3_RF.fa		gi|34067|emb|X032 12.1|HSKER7R	x
TCC-70E7	none:17_TCC_70E7_1_M13F.fa none:17_TCC_70E7_1_M13R.fa			x
TCC-21G7	Homo sapiens clone PP722 unknown mRNA; CONTIG_nt_non_genomic(identity):contig_TCC_21G7_RF.fa		gi|10441985|gb|AF2 18028.1|AF218028	
TCC-93G5	Homo sapiens cystatin B (CSTB) gene, promoter reg ; nt_non_genomic(identity):contig_TCC_93G5_RF.fa		gi|7263011|gb|AF20 8234.1|AF208234	x
TCC-36B5	Sequence 5 from Patent WO9954447 ; nt_non_genomic(identity):contig_TCC_36B5_RF.fa	Homo sapiens hypothetical protein (LOC51323), mRNA	gi|10040588|emb|A X014141.1|AX0141 41	
TCC-54C11	Homo sapiens actin, gamma 1 (ACTG1) mRNA ; nt_non_genomic(identity):contig_TCC_54C11_RF.fa		gi|4501886|ref|NM_ 001614.1|	
TCC-34A5	Homo sapiens S100 calcium-binding protein P (S100P ; nt_non_genomic(identity):contig_TCC_34A5_RF.fa		gi|5174662|ref|NM_ 005980.1|	x
TCC-70E8	none:contig_TCC_70E8_RF.fa			x
TCC-78B11	Sequence 56 from Patent WO9954353 ; nt_non_genomic(identity):contig_TCC_78B11_RF.fa	Human growth factor-inducible 2A9 gene, complete cds		
TCC-101E11	Homo sapiens CGi-81 protein (LOC51108), mRNA ; nt_non_genomic(identity):contig_TCC_101E11_RF.fa		gi|7705788|ref|NM_ 016025.1|	x
TCC-102C5	MR1-CT0058-021199-001-c10 CT0058 Homo sapiens cDNA ; est(identity):contig_TCC_102C5_RF.fa		gi|6879340|gb|AW3 74686.1|AW374686	
TCC-58A3	Homo sapiens keratin 17 (KRT17) mRNA; nt_non_genomic(identity):contig_TCC_58A3_RF.fa		gi|30378|emb|Z195 74.1|HSCYTOK17	
TCC-57B3	Homo sapiens solute carrier family 2 (facilitated ; nt_non_genomic(identity):contig_TCC_57B3_RF.fa		gi|5730050|ref|NM_ 006516.1|	x
TCC-42G5	Homo sapiens caspase 4, apoptosis-related cystein ; nt_non_genomic(identity):contig_TCC_42G5_RF.fa		gi|4502576|ref|NM_ 001225.1|	
TCC-99G12	Homo sapiens keratin 8 (KRT8) mRNA ; nt_non_genomic(identity):contig_TCC_99G12_RF.fa		gi|4504918|ref|NM_ 002273.1|	
TCC-92D7	Homo sapiens hypothetical protein PRO2987 (PRO298 ; nt_non_genomic(identity):contig_TCC_92D7_RF.fa		gi|8924228|ref|NM_ 018636.1|	
TCC-89G3	Sequence 82 from Patent WO9951727 ; nt_non_genomic(identity):contig_TCC_89G3_RF.fa	Homo sapiens midkine (neurite growth-promoting factor 2) (MDK) mRNA	gi|10041391|emb|A X015411.1|AX0154 11	x
TCC-56E11	Homo sapiens Opa-interacting protein OIP3 mRNA, p ; nt_non_genomic(identity):contig_TCC_56E11_RF.fa		gi|2815605|gb|AF02 5439.1|AF025439	x
TCC-25F2	Sequence 89 from Patent WO9953040 ; nt_non_genomic(identity):contig_TCC_25F2_RF.fa	Homo sapiens fatty acid binding protein 5 (psoriasis-associated) (FABP5), mRNA	gi|10041157|emb|A X014890.1|AX0148 90	x
TCC-44C1	Homo sapiens S100 calcium-binding protein A13 (S10 ; nt_non_genomic(identity):contig_TCC_44C1_RF.fa		gi|5174658|ref|NM_ 005979.1|	x

Notes to Table 5

1) In column A ,"x" indicates that the sequence also appears in Table 1 or 2.

2) Table 5 includes known genes whose function in bladder cancer was

heretofore unknown and which were now found to upregulated in bladder cancer (identified by Accession Number) and also includes sequences of novel genes which have no identity to known proteins or genes in the gene databases

5

10

15

20

25

30

35

40

45

Table 6 Polynucleo...

```
>17_TCC_70E7_1_M13F.fa
GGTAGACGTACCTGCGTCCCAGACTTGACCAGGTGGATCTCCTGTTTTAC
TCACGAGGACTTTCCCAGGAAAACCATGCCACTAGCAAAATAATATAAAC
AAAGGA
>17_TCC_70E7_1_M13R.fa
TTTTTTTTTTTTTTTTTGGCTAGAGGCATGGATATCCTGGGAAAGCTCTCC
TGAGTAAAAGACGAGAGACACCTGGTGAAGACTGGAACGCATGTACGTCT
ACC
>contig_TCC_101E11_RF.fa
GGTCGACGTACCTGCGCAATAAAGCTGTCCATTCAATTCCAAATACTGGT
TTTAAGGTATAGCCACTGATATTCTTTCATGTTTAGAAATTCTTTCTGTT
ATTATTCAAGAAAATGTTTTTAATCATGCTAATAAACTTTTTTGGAGATG
AAAAAAAAAAAAAAAAAAA
>contig_TCC_102D5_RF.fa
GCTGGTTGGGGGAATTGGAGGCTTCTAGGAGGTGGCACGGTGCACGCCAA
GATGGCTGTGTCCACAGAGGAGCTGGAGGCCACGGTTCAGGAAGTCCTGG
GGAGACTGAAGAGCCACCAGTTTTTCCAGTCCACATGGGACACTGTTGCC
TTCATTGTTTTCCTCACCTTCATGGGCACCGTGCTGCTCCTGCTGCTGCT
GGTCGTCGCCCACTGCTGCTGCTGCAGCTCCCCCGGGCCCCGCAGGGAAA
GCCCCAGGAAGGAAAGACCCAAGGGAGTGGATAACTTGGCCCTGGAACCC
TGACCCTGTGTCTCCTGCCCGGTGGCAGTAACAAAGCCTTCTGTCTGCCC
AGAAAAAAAAAAAAAAAAA
>contig_TCC_21G7_RF.fa
CTAAATCTAGGTATTCTGGCTGAGTGTATCTGGGTGGGCCAGCTAAAAAT
AAACCTCATTGAACTCCAGCCCAACCCAGAGAAACATCCAGAAGAGCCT
TGAATTAGTGATCCAAAACCCAGGGGGAAAGGCGACATTCTCACCCCCAG
CACCCCCTTCACCTCACCTCAACTCCTACTCTCTCGGTCTATAATCACTG
CTCTCTCTCTCCCCAACACCACTATTGAACAGGAGCCCTTGTCACCAGGT
CCAAGCAATTCCCTAAGGTATCACAAACAATGGTGGATGCAATTTTACCT
TACTCAGTAACCACGAGGCTCACATCCCTAATTTCAGACTCTACCAGCTC
TCAGGTGCCCTCCCAAGGGGCTGCCTGCATGAAGATGCCTTGGAAGTAGC
CCCTTTCACAATCACAGGAATTAACCCCCTGGTGTTGGAGGGGCCTCACT
TTAAGCAATCCCAGTAGTAAACATTGGATAAATCTAAAGGCTTTCTTTAA
TTTTTTTTTTCTCTTCGTAAAGGATTCAAAGCAGGCACAGTGGTG
>contig_TCC_25F2_RF.fa
CCCTGGGAGAGAAGTTTGAAGAAACCACAGCTGATGGCAGAAAAACTCAG
ACTGTCTGCAACTTTACAGATGGTGCATTGGTTCAGCATCAGGAGTGGGA
TGGGAAGGAAAGCACAATAACAAGAAAATTGAAAGATGGGAAATTAGTGG
TGGAGTGTGTCATGAACAATGTCACCTG
>contig_TCC_34A5_RF.fa
CATGAGCAGGCTCAGCCTAGGGGAATAATTGCCAACAAACACTTTTGGGA
AGCCTGGGACCATGGCTCTGCCAGGAATCTGTGACATCTCCAGGGCATCA
TTTGAGTCCTGCCTTCTCAAAG
>contig_TCC_36B5_RF.fa
CTCTTCTTATGCTAATATGCTCTGGGCTGGAGAAATGAAATCCTCAAGCC
ATCAGGATTTGCTATTTAAGTGGCTTGACAACTGGGCCACCAAAGAACTT
GAACTTCACCTTTTAGGATTTGAGCTGTTCTGGAACACATTGCTGCACTT
TGGAAAGTCAAAATCAAGTGCCGTGGCGCCCTTTCCATAGAGAATTTGC
CCAGCTTTGCTTTAAAAGATGTCTTGTTTTTTATATACACATAATCAATA
GGTCCAATCTGCTCTCAAGGCCTTGGTCCTGGTGGGATTCCTTCACCAAT
```

97

```
TACTTTAATTAAAAATGG......
TGCAACTATGCTCCCATTTACAAATG
>contig_TCC_42G5_RF.fa
CCTTCCGAAATACTTCCTCCAGGTGGCAGCACCAAGAATATTTCTGGAAG
CATGTGATCAGTTGTGTGATGAAGATAGAGCCCATTGTGCTGTCTCTCCA
GGACACGTTGTGTGGCGTTGAAGAGCAGAAAGCAATGAAGTCCTTCTCCA
CGTGGGTCTTGTAAACAGCATCTTCCTCCAGGTTCTCAGATGACTGTGAA
GAGGCCACTTCCAAGGATGCTGGAGAGTCTCTGACCCACAGTTCCCCACG
GTTTGCACCTCTGCAGGCCTGGACAATGATGACCTTGGGTTTGTCCTTCA
GACTGAGGCAGTTGCGGTTGTTGAATATCTGGAAGATGGTGTCATAAAGC
AGCACATCTGGTTTTTTCTCATCATGCACAGTTCCGCAGATTCCCTCCAG
GATGCCATGAGACATGGG


>contig_TCC_44C1_RF.fa
CTCATTGAACTTGAGCTCCGAGTCCTGATTCACATCCAAGCTCTTCATCT
TCTCATCAAGAGAGCCCACATCCTTGAGCAGATGGGGCAACTGCTGGGTA
ACCAGCTCTTTGAACTCGTTGACGCTGAGGCTATCCTTCCGGCCCTCCTG
CCTTGCAAAGGTGAAGAAGGTGGTGACCACGACCCTGAGGCCAAAGCTGATG
GCTCTGTCAGTGGTTCTGCTGCCATTAGGGCTGACCTCTCCTCAGCGGC
TCCTCAAGGGGCTAGCTGACCTTTGTCAGGGCTGACCTCTCCTCAGCGGC
AGCAGGGCAGAGTGCTGAACCCAGGACCCCACAGATCCTCCCCGCTCCTG
TCTCCCGGTGACAAGGGTCCTGGAACGGGGCGTCTCTGACTCCCTGCTCC
AGGACGGGTTTAG
>contig_TCC_54C11_RF.fa
TTTTTTTTTTTTTTTTTTGGTTACGGCAGCACTTTTATTTTTCCTTACACA
ATGACGTGTTGCTGGGGCCTAATGTTCTCACATAACAGTAGAAAACCAAA
ATTTGTTGTCATCTCTTCAAAGAATCGAGAATTGCGTACAAAAAAAAAAA
AAAAAAA
>contig_TCC_56E11_RF.fa
CTCTCCAGTTTGCACCTGTCCCCACCCTCCACTCAGCTGTCCTGCAGCAA
ACACTCCACCCTCCACCTTCCATTTTCCCCCACTACTGCAGCACCTCCAG
GCCTGTTGCTATAGAGCCTACCTGTATGTCAATAAACAACAGCTGAAGCA
AAAAAAAAAAAAAAA
>contig_TCC_57B3_RF.fa
GGTACGACGGACCTGCGGAGACTCCTGCCCTGTTGTGTATAGATGCAAGA
TATTTATATATATTTTTGGTTGTCAATATTAAATACAGACACTAAGTTAT
AGTATATCTGGACAAGCCAACTTGTAAATACACCACCTCACTCCTGTTAC
TTACCTAAACAGATATAAATGGCTGGTTTTTAGAAAAAAAAAAAAAAAAAA
A
>contig_TCC_58A3_RF.fa
GGCTGGAGCAGGAGATTGCCACCTACCGCCGCCTGCTGGAGGGAGAGGAT
GCCCACCTGACTCAGTACAAGAAAGAACCGGTGACCACCCGTCAGGTGCG
TACCATTGTGGAAGAGGTCCAGGATGGCAAGGTCATCTCCTCCCGCGAGC
AGGTCCACCAGACCACCCGCTGAGGACTCAGCTACCCCGGCCGGCCACCC
AGGAGGCAGGGAGGCAGCCGCCCCATCTGCCCCACAGTCTCCGGCCTCTC
CAGCCTCAGCCCCCTGCTTCAGTCCCTTCCCCATGCTTCCTTGCCTGATG
ACAATAAAGCTTGTTGACTCAGCTAAAAAAAAAAAAAAAAAAA

>contig_TCC_70E8_RF.fa
TTTTTTTTTTTTTTTTTGCTAGTGGCATGGTTTTCCTGGGAAAGTCCTCC
TGAGTAAAAGAGGAGAGACACCTGGTGAAGACTGGGACGCAGGTACGTCT
```

                                                          98

ACT
>contig_TCC_71E4_RF.fa
CTCCTGCGATATGTTCAACTATGAAGAATACTGCACCGCCAACGCAGTCA
CTGGGGCTTGCCGTGCATCCTTCCCACGCTGGTACTTTGACGTGGAGAGG
AACTCCTGCAATAACTTCATCTATGGAGGCTGCCGGGGCAATAAGAACAG
CTACCGCTCTGAGGAGGCCTGCATGCTCCGCTGCTTCCGGCGGGGCTGTTC
ATCCTCCGCTGCCCCTTGGCTCAAAGGTGGTGGTTCTGGCGGGGCTGTTC
GTGATGGTGTTGATCCTCTTCCTGGGAGCCTCCATGGTCTACCTGATCCG
GGTGGCACGGAGGAACCAGGAGCGTGCCCTGCGCACCGTCTGGAGCTCCG
GAGATGACAAGGAGCAGCTGGTGAAGAACACATATGTCCTGTGACCGCCC
TGTCGCCAACAGGACTGGGAAGGGAGGGGAGACTATGTGTGAGCTTTTTT
TAAATAGAGGGATTGACTCGGATTTGAGTGATCATTAGGGCTGAGGTCTG
TTTCTCTGGGAGGTAGGACGGCTGCTTCCTGGTCTGGCAGGGATGGGTTT
GCTTTGGAAATCCTCTAGGAGGCTCCTCCTCGCATGGCCTGCAGTCTGGC
AGCAGCCCCGAGTTGTTTCCTCGCTGATCGATTTCTTTCTCATCACAGAAG
GTTTTCTTTGCTTATGTTGAATTCCATTGCCTCTTTTCTCATCACAGAAG
TGATGTTGGAATCGTTTCTTTTGTTTGTCTGATTTATGGTTTTTTTTAAGT
ATAAACAAAAGTTTTTTATTAGCATTCTGAAAGAAGGAAAGTAAAATGTA
CAAGTTTAATAAAAAGGGGCCTTCCCCTTTAGAATAAATTTCAGCATGTG
CTTTCAAAAAAAAAAAAAAAAAAAA
>contig_TCC_75E3_RF.fa
AAAGAGGGCGGCAGGGCCTGGAGATCCTCCTGCAGACCACGCCCGTCCT
GCCTGTGGCGCCGTCTCCAGGGGCTGCTTCCTCCTGGAAATTGACGAGGG
GTGTCTTGGGCAGAGCTGGCTCTGAGCGCCTCCATCCAAGGCCAGGTTCT
CCGTTAGCTCCTGTGGCCCACCCTGGGCCCCTGGGCTGGAATCAGGAATA
TTTTCCAAAGAGTGATAGTCTTTTGCTTTTGGCAAAACTCTACTTAATCC
AATGGGTTTTTCTCTGTACAGTAGATTTTCCAAATGTAATAAACTTTAAT
ATAAAGTAAAAAAAAAAAAAAAAAAA
>contig_TCC_78B11_RF.fa
GGACCGGAACAAGGACCAGGAGGTGAACTTCCAGGAGTATGTCACCTTCC
TGGGGGCCTTGGCTTTGATCTACAATGAAGCCCTCAAGGGCTGAAAATAA
ATAGGGAAGATGGAGACACCCTCTGGGGGTCCTCTCTGAGTCAAATCCAG
TGGTGGGTAATTGTACAATAAATTTTTTTTGGTCAAATTTAAAAAAAAAA
AAAAAA
>contig_TCC_89G3_RF.fa
CAGGAGACCATCCGCGTCACCAAGCCCTGCACCCCCAAGACCAAAGCAAA
GGCCAAAGCCAAGAAAGGGAAGGGAAAGGACTAGACGCCAAGCCTGGATG
CCAAGGAGCCCCTGGTGTCACATGGGGCCTGGCCCACGCCCTCCCTCTCC
CAGGCCCGAGATGTGACCCACCAGTGCCTTCTGTCTGCTCGTTAGCTTTA
ATCAATCATGCCCTGCCTTGTCCCTCTCACTCCCCAGCCCCACCCCTAAG
TGCCCAAAGTGGGGAGGGGACAAGGGATTCTGGGAAGCTTGAGCCTCCCCC
AAAGCAATGTGAGTCCCAGAGCCCGCTTTTGTTCTTCCCCACAATTCCAT
TACTAAGAAACACATCAAATAAACTGACTTTTTCCCCCCAAAAAAAAAAA
AAAAA
>contig_TCC_92D7_RF.fa
TTTTTTTTTTTTTTGAAGACAACTTTTACAAACTGATGTTTATTTTCCA
TCAACCATTTTTCCATGCTGCTTAAGAGCCTATGCAAGAACAGCTTAAGA
CCAGTCAGTGGTTGAAGTC
>contig_TCC_93G5_RF.fa
GACTACCAGACCAACAAAGCCAAGCATGATGAGCTGACCTATTTCTGATC
CTGACTTTGGACAAGGCCCTTCAGCCAGAAGACTGACAAAGTCATCCTCC
GTCTACCAGAGCGTGCACTTGTGATCCTAAAATAAGCTTCATCTCCGGGC

99

```
TGTGCCCCTTGGGGTGGA---------
CTCTTCTAAATTTCATTGTGTTGATTTCTTTCCTTCCCAATAGGTGATC
TTAATTACTTTCAGAATATTTTCAAAATAGATATATTTTTAAAATCCTTA
CAAAAAAAAAAAAAAAA
>contig_TCC_94G3_RF.fa
AAGGCTTATTCCATCCGGACCGCATCCGCCAGTCGCAGGAGTGCCCGCGA
CTGAGCCGCCTCCCACCACTCCACTCCTCCAGCCACCACCCACAATCACA
AGAAGATTCCCACCCCTGCCTCCCATGCCTGGTCCCAAGACAGTGAGACA
GTCTGGAAAGTGATGTCAGAATAGCTTCCAATAAAGCAGCCTCATTCTGA
GGCCTGAGTGAAAAAAAAAAAAAAAAAAAA
>contig_TCC_99G12_RF.fa
AGCGGCTATGCAGGTGGTCTGAGCTCGGCCTATGGGGGCCTCACAAGCCC
CGGCCTCAGCTACAGCCTGGGCTCCAGCTTTGGCTCTGGCGCGGGCTCCA
GCTCCTTCAGCCGCACCAGCTCCTCCAGGGCCGTGGTTGTGAAGAAGATC
GAGACACGTGATGGGAAGCTGGTGTCTGAGTCCTCTGACGTCCTGCCCAA
GTGAACAGCTGCGGCAGCCCCTCCCAGCCTACCCCTCCTGCGCTGCCCCA
GAGCCTGGGAAGGAGGCCGCTATGCAGGGTAGCACTGGGAACAGGAGACC
CACCTGAGGCTCAGCCCTAGCCCTCAGCCCACCTGGGGAGTTTACTACCT
GGGGACCCCCCTTCCCCATGCCTCCAGCTACAAAACAATTCAATTGCTTT
TTTTTTTTTGGTCCAAAATAAAACCTCAGCTAGCTCTGCCAATGTCAAAA
AAAAAAAAAAAAAAAA
```

The first two sequences are from opposite ends of the same polynucleotide, (and are thus in the same gene). All the other 21 sequences are contigs.

100

AMERICAN CANCER SOCIETY, Estimating new cancer cases and deaths by sex for all sites, Cancer Facts and Figures, pg. 4 (1998)

Burke and Olson, "Preparation of Clone Libraries in Yeast Artificial-Chromosome Vectors" in Methods in Enzymology, Vol. 194, "Guide to Yeast Genetics and Molecular Biology", eds. C. Guthrie and G. Fink, Academic Press, Inc., Chap. 17, pp. 251-270 (1991).

Capecchi, "Altering the genome by homologous recombination" Science 244:1288-1292 (1989).

Cregg JM, Vedvick TS, Raschke WC: Recent Advances in the Expression of Foreign Genes in Pichia pastoris, Bio/Technology 11:905-910, 1993

Culver, 1998. Site-Directed recombination for repair of mutations in the human ADA gene. (Abstract) Antisense DNA & RNA based therapeutics, February, 1998, Coronado, CA.

Davies et al., "Targeted alterations in yeast artificial chromosomes for inter-species gene transfer", Nucleic Acids Research, Vol. 20, No. 11, pp. 2693-2698 (1992).

Diatchenko et al., "Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries" Proc. Nat Acad. Sci., Vol. 93, pp. 6025-6030 (1996):

101

Dickinson et al., ... ~ . ~ .   na!

vectors", <u>Human Molecular Genetics</u>, Vol. 2, No. 8, pp. 1299-1302 (1993).

Duff and Lincoln, "Insertion of a pathogenic mutation into a yeast artificial chromosome containing the human APP gene and expression in ES cells", <u>Research Advances in Alzheimer's Disease and Related Disorders</u>, 1995.

Gilboa, E, Eglitis, MA, Kantoff, PW, Anderson, WF: Transfer and expression of cloned genes using retroviral vectors. BioTechniques 4(6):504-512, 1986.

GROSSMAN, HB, Semin. Urol. Oncol., 16(1):17-22 (1998)

HUDSON MA, and HERR HW, Carcinoma *in situ* of the bladder, J. Urol., 153(3, Part 1):564-572 (1995)

HUDSON, MA and HERR HW, Carcinoma *in situ* of the bladder, J. Urol., 153(3) Part 1):564-572 (1995)

Huston et al, 1991 "Protein engineering of single-chain Fv analogs and fusion proteins" in Methods in Enzymology (JJ Langone, ed.; Academic Press, New York, NY) 203:46-88.

Huxley et al., "The human HPRT gene on a yeast artificial chromosome is functional when transferred to mouse cells by cell fusion", <u>Genomics</u>, 9:742-750 (1991).

Jakobovits et al., "Germ-line transmission and expression of a human-derived yeast artificial chromosome", <u>Nature</u>, Vol. 362, pp. 255-261 (1993).

102

:

Johnson and ~~~~, derivatives of monoclonal antibodies and their production in *Escherichia coli* in Methods in Enzymology (JJ Langone, ed.; Academic Press, New York. NY) 203:88-99.

LACOMBE L et al., Overexpression of p53 protein in a high-risk population of patients with superficial bladder cancer before and after bacillus Calmette-Guerin therapy; correlation to clinical outcome, J. Urol., 153(3, Part 1):564-572 (1995)

Lamb et al., "Introduction and expression of the 400 kilobase precursor amyloid protein gene in transgenic mice", <u>Nature Genetics</u>, Vol. 5, pp. 22-29 (1993).

Mernaugh and Mernaugh, 1995 "An overview of phage-displayed recombinant antibodies" in Molecular Methods In Plant Pathology (RP Singh and US Singh, eds.; CRC Press Inc., Boca Raton, FL) pp. 359-365.

OZEN H., Curr. Opin. Oncol., 10(3):273-8 (1998)

Pearson and Choi, *Expression of the human β-amyloid precursor protein gene from a heast artificial chromosome in transgenic mice.* Proc. Natl. Scad. Sci. USA, 1993. 90:10578-82.

Rothstein, "Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast" in <u>Methods in Enzymology</u>, Vol. 194, "Guide to Yeast Genetics and Molecular Biology", eds. C. Guthrie and G. Fink, Academic Press, Inc., Chap. 19, pp. 281-301 (1991).

Schedl et al., ...,

tyrosinase gene confers copy number-dependent expression in transgenic mice".

Nature, Vol. 362, pp. 258-261 (1993).

Strauss et al., "Germ line transmission of a yeast artificial

chromosome spanning the murine α_1 (I) collagen locus", Science, Vol. 259, pp.

1904-1907 (1993).

TORTI FM and LURN BL, The biology and treatment of superficial

bladder cancer, J. Clin. Oncol., 2(5):505-531 (1984)